

Assessing Moral Development in the Liberal Arts

Author: Kerry Cronin

Persistent link: <http://hdl.handle.net/2345/bc-ir:104138>

This work is posted on [eScholarship@BC](#),
Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2015

Copyright is held by the author, with all rights reserved, unless otherwise noted.

Boston College
Lynch School of Education

Department of
Educational Administration and Higher Education
Higher Education Program

ASSESSING MORAL DEVELOPMENT
IN THE LIBERAL ARTS

Dissertation
by
KERRY M. CRONIN

submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

May, 2015

© Copyright by Kerry M. Cronin
2015

ASSESSING MORAL DEVELOPMENT IN THE LIBERAL ARTS

Kerry M. Cronin

Boston College Lynch School of Education

Dissertation Chair: Karen Arnold

Liberal education has long claimed moral education to be a chief aim of its educational format. Liberal education supporters regularly assert its unique ability to foster moral and ethical development in students, but data regarding higher education's efficacy in promoting moral development are limited. Additionally, the educational goal of moral development suffers important philosophical and epistemological critiques which bring into question its adequacy as a worthwhile aim of contemporary higher education. In order to discern whether higher education resources should be used to pursue this educational objective, liberal arts practitioners and supporters must identify clearly what moral education is, whether it is a facet of college student development worthy of our attention, and how to adequately measure it. This study offers a careful analysis of data related to student moral reasoning development gathered in an evaluation process of a liberal education course at a mid-sized research institution. The central research questions focus on aspects of student moral development and students' perceptions of the moral dimensions of coursework and highlight how these interact with students' abilities to receive and process course materials and activities. The research design employs a concurrent triangulation approach to quantitative and qualitative course assessment materials. James Rest's Defining Issues Test (DIT), a well-researched, neo-Kohlbergian measure of moral reasoning, and student writing were analyzed in pre- and

post-course evaluations to investigate students' moral reasoning development as they entered, changed and left a year-long liberal arts course. Results reveal important features of student moral growth, illuminating how students at different levels of moral reasoning development and with varying degrees of change with respect to moral reasoning engaged with liberal education course materials and activities in quite distinct ways. This is an important step in uncovering the unique aspects of liberal education that may foster and sustain moral growth.

Acknowledgements

Though it seems that most of the time spent thinking about and writing a dissertation is solitary work, it has been very clear to me that I have not been alone for much of it at all – one way or another. Many wonderful people have carried me through the long days, weeks and years of this project and fortified me with all sorts of necessary provisions for the journey, from aid in figuring out just the right words or journal articles to read, to patience and prayers when I all but disappeared to read, write, or analyze formulas. It's hard to capture in a brief note the gratitude I have felt steadily over these many years.

The range of my director's mentoring abilities is astounding. The handprint of Karen Arnold's wisdom, patience and eagle-eye editing is on every page of this work. When I asked her to direct my dissertation, we joked that the process of "the ask" feels a lot like asking someone on a date and now I cannot imagine a better dance partner. Karen has an uncanny ability to sense what will really work at any given time. One the one hand, she will sit patiently listening to a student ramble on and on about a topic and then after pondering for less than a minute, will summarize the central question with the ease of a dissertation Ninja. She seems to have a sixth sense about what is needed and when it is needed: ample space or a gentle nudge, a critical eye or a cheerleading session, a discussion about stats or just perhaps about religious ideas or daily practices. One of the most startling gifts I received in this process was being treated by Karen as a research colleague throughout my program of study. Her intellectual and personal generosity to

her students is unflagging and my teaching has improved as I emulate the consistent respect and support she shows her students.

I am lucky to have had two other master teachers reading this dissertation: Audrey Friedman and Ronald Tacelli, who both replied affirmatively to my request to serve on the committee within minutes of my asking. Their generosity and support along the way have been unwavering. Each brings a particular area of expertise to the study of moral development but for these two scholars, the moral life is not simply an intellectual exercise but rather a way of life that is shared generously. I have also had the great fortune of having four supervisors in my professional life who have all been master teachers and amazing individuals. Sr. Patricia Andrews, Joseph Flanagan, SJ, Dr. Patrick Byrne and Dr. Brian Braman have each taught me much more about life and learning than one would expect to discover while making a living. I have been profoundly impacted by each of them personally and professionally.

It would be impossible to make a complete list the family and friends who have supported me through this program and dissertation. Their love and support came in many formats over the years: prayers, food, a well-timed glass of wine or “pink drink,” a message of support with no demand for a return call, laughter during long evenings studying for the comps, extra help and a knowing look at the office, intuition about when to ask about my writing and when not to, patience with endless delays, a shoulder on which to lean or cry. My parents, Timothy and Rosemarie Cronin, seem to think I am “the best thing since sliced bread” and they have made me feel loved every day of my life. I love them fiercely and I sincerely hope that my work honors them. Everything that is even remotely good about me can be traced back to my family and the love that has

enveloped me throughout my life. My brothers, sisters-in-law, nieces and nephews, as well as a slew of cousins, aunts and uncles have been devoted supporters of me on this journey and I am forever grateful for their humor, compassion and encouragement along the way. Three of my very dearest uncles passed away during the writing of this dissertation but I have felt in very real ways the presence and love of these great men, Joseph Flanagan, SJ, Rev. James Cronin, and Kevin Flanagan, every step of the way. I dedicate this work to their memory.

Friends who've supported me over the years are too numerous to count but too remarkable to gloss over. My life is overfilled with genuine friends who inspire and challenge me daily. I realized in the course of writing this dissertation that much of what I understand about moral goodness comes from those friends who show me day in and day out what goodness really means and how much fruit it bears in the lives of others. It is a great gift indeed to receive love given freely by a friend and I am deeply and profoundly grateful for the many wonderful friends in my life. It is appropriate to include in this paragraph about friendship a special note of gratitude to the hundreds of students I have had the privilege of teaching throughout my career. When I began teaching so many years ago I never expected how much sheer joy students would bring into my life.

I would like to add a note of thanks to Boston College, particularly the Lonergan community, for being a second home to me in so many ways over many years. I am especially grateful to the faculty and staff of the Higher Education Department in the Boston College Lynch School of Education for their patient mentoring, steadfast support and sincere collegiality.

Finally, I believe that this dissertation is only successful to the extent that it makes more intelligible the work that God accomplishes in each of us. Moral goodness is a mark of human freedom and dignity and these are some of the greatest gifts that God bestows on each of us. I am grateful for all of the many blessings I have received.

Table of Contents

Acknowledgments	i
Table of Contents	v
List of Tables	x
Chapter 1: Moral Education in American Higher Education	1
1.1 Introduction	1
1.2 Moral Education's History in American Higher Education	3
1.2.A. Moral Education within Great Books Programs	8
1.3 The Contemporary Debate over Moral Education in the Academy	10
1.3.A Contemporary Research and Moral Education	15
1.4 The Aims of the Present Study	18
Chapter 2: Literature Review	22
2.1 Moral Development Research	23
2.2 Lawrence Kohlberg's Work and Influence	27
2.3 Neo-Kohlbergian Research: James Rest and the Minnesota Group	31
2.4 Three Neo-Kohlbergian Research Directions	33
2.4.A. Development of the DIT	34
2.4.B. Schema Theory and the DIT Schemas	36
2.4.C. The Four Component Model	41
2.5 A Neo-Kohlbergian Definition of Morality	43
2.6 Kohlbergian to Neo-Kohlbergian Research: Development and Critiques	44
2.6.A. Critiques of Foundationalism	45
2.6.B. Gender Matters	47

2.6.C. Critiques of the DIT as a Measurement Tool: What Does It Measure?	50
2.7 Recent Research with the DIT: Educational and Pedagogical Interventions in Moral Development Research	54
2.8 Implications for Liberal Education and the Boston College Perspectives Program	60
Chapter 3: Research Intervention and Quantitative Methodology	63
3.1 Introduction: rationale of the research questions and design	63
3.2 The Research Questions	65
3.3 The Research Design	65
3.4 The Research Site and Treatment/Intervention	69
3.4.A The Research Site: Boston College	69
3.4.B The Treatment/Intervention and Sample: “Perspectives in Western Culture”	72
3.5 Data, Data Collection and Sample	74
3.5.A Data	74
3.5.B Data Collection	75
3.5.C The Sample	77
3.6 The Quantitative Instrument: The DIT	78
3.6.A History and Development of the DIT	78
3.6.B The DIT in Detail	80
3.6.C Scoring the DIT: From P Scores to the N2 Index	82
3.6.D Calculating N2 scores	84
3.6.E DIT2 Validity and Reliability	86

3.7 Conclusion	93
Chapter 4: Qualitative/Rubric Methodology	95
4.1 Introduction: Rationale for Qualitative Analysis of Writing	95
4.2 The Research Question	96
4.3 The Research Design	97
4.4 Data and Data Collection	99
4.4.A Subsample Essay Selection Process	100
4.4.A.1 Selection of Gainer Essays	102
4.4.A.2 Selection of Decliner Essays	102
4.4.A.3 Selection of High to High Essays	103
4.4.A.4 Selection of Low to Low Essays and New Gainer Group	103
4.5 Design of Rubrics and Rubric Items	105
4.5.A Rubric Item Development	105
4.5.B Rubric Items from DIT studies	106
4.5.C Rubric Items from Meta-analyses and Survey Benchmarking	107
4.5.D Piloting of Rubric	108
4.6. Rubrics	111
4.7 Data Analysis - Rubric Application	118
4.8 Limitations of the Study	120
Chapter 5: Analysis of Quantitative Data and Findings	125
5.1 Introduction	125
5.2 DIT Sample and DIT Analysis Procedure	126
5.3 DIT Findings	127

5.3.A Sample N2 Mean Scores and National Norms	127
5.3.B Sample N2 Change Scores	131
5.4 Groupings of DIT Scores	132
5.4.A DIT Groups from DIT analysis	134
5.4.A.1 DIT Gainers	135
5.4.A.2 DIT Decliners	136
5.4.A.3 DIT High to High Scorers	137
5.4.A.4 DIT Low to Low Scorers	138
5.5 Findings across groups	139
Chapter 6: Analysis of Qualitative Rubric Data and Findings	145
6.1 Introduction	145
6.2 Essay Subsample	146
6.3 Essay Analysis	148
6.4 Essay Analysis: Rubric Analyses	148
6.4.A Rubric Analysis by Group	149
6.4.A.1 Gainers' Rubric Scores	149
6.4.A.2 Decliners' Rubric Scores	154
6.4.A.3 High to High Scorers' Rubric Scores	160
6.4.A.4 Low Gainers' Rubric Scores	163
6.4.2 Rubric Patterns across Groups	169
Chapter 7: Essay Coding Analyses and Findings	176
7.1 Essay Analysis: Themes/Motifs Detected in Open and Axial Coding	176
7.1.A Themes within Groups	176

7.1.A.1 Gainers' Themes	176
7.1.A.2. Decliners' Themes	179
7.1.A.3. High to High Scorers' Themes	186
7.1.A.4. Low Gainers' Themes	192
7.2 Essay Analysis: Themes/Motifs Detected in Selective Coding	196
7.2.A. Moral Development Themes across Groups	197
7.2.B Course-Related Themes across Groups	203
7.3 Essay Analyses: Limitations of Findings	206
7.4 Essay Analyses: Conclusion	208
Chapter 8: Implications of the Study and Conclusion	210
8.1 Review of the Research Questions and Design	210
8.2 Summary of the Study Findings	212
8.3 Interpretation of the Study Findings	217
8.3.A Interpretation of DIT Findings	217
8.3.B Interpretation of Writing Analyses	222
8.3.C Limitations of the Study	225
8.4 Implications for the Course, Perspectives in Western Culture	227
8.4.A Effective Practices of the Course	233
8.5 Implications for Liberal Education Program Design and Implementation	235
References	240

List of Tables

Table 2.1	Kohlberg's Stages vs. Rest's Schemas	40
Table 3.1	Concurrent Triangulation Design	69
Table 4.1	Writing Rubric Items, Moral Development Proxies	114
Table 4.2	Writing Rubric Items, Course-related Items for Coding	116
Table 5.1	Study Sample N2 Pre- and Post-test scores	127
Table 5.2	Comparison of N2 Scores and National Norms	128
Table 5.3	Comparisons of Pre-test N2 Scores of Portico and Perspectives	130
Table 5.4	Descriptive Statistics of N2 Changes	131
Table 5.5	T-test Statistics, Correlations and Paired Differences	132
Table 5.6	Gainers' Descriptive Statistics	135
Table 5.7	Decliners' Descriptive Statistics	136
Table 5.8	High to High Scorers' Descriptive Statistics	138
Table 5.9	Low Gainers' Descriptive Statistics	139
Table 5.10	Group Statistics for Total Sample	140
Table 5.11	Quartile Analysis Results	141
Table 5.12	Scatterplot of Subsample Decliners	143
Table 6.1	Essay Subjects from Two-class Subsample	146
Table 6.2	Full Sample Quartile Analysis	147
Table 6.3	Gainers' Moral Development Rubric Outcomes	151
Table 6.4	Gainers' Moral Course-related Rubric Outcomes	153
Table 6.5	Decliners' Moral Development Rubric Outcomes	155
Table 6.6	Decliners' Moral Course-related Rubric Outcomes	159

Table 6.7	High to High Scorers' Moral Development Rubric Outcomes	161
Table 6.8	High to High Scorers' Course-related Rubric Outcomes	163
Table 6.9	Low Gainers' Moral Development Rubric Outcomes	165
Table 6.10	Low Gainers' Course-related Rubric Outcomes	168
Table 6.11	Moral Development Rubric Items Ranked by Group	170
Table 6.12	Moral Development Rubric Items Listed Alphabetically with Group Scores	172
Table 6.13	Course-related Rubric Items Rankings, Highest to Lowest	174

Chapter 1:

Moral Education in American Higher Education

1.1 INTRODUCTION

Among the many goals and purposes of American higher education, the aim of advancing college students' moral or character development is positioned among the most contentious in the contemporary scene. Recent conversations among politicians and leaders in higher education have recommenced the public debate about the fundamental aims of American higher education, opening the door to a renewed consideration of this once preeminent but now embattled educational aspiration. In the past two decades, a number of higher education leaders have offered defenses of the aim of moral education, recommending serious reflection on what is lost when higher education jettisons this goal (Arum & Roksa, 2010; Bok, 2007; Delbanco, 2012; Gutmann, 1997; King, 1997; Kronman, 2008; Lewis, 2006; Reuben, 1996). A concurrent conversation in the political sector has emerged in the past few years regarding these same sorts of questions but from a different vantage point. This conversation is marked by calls for reform and oversight of institutions to make postsecondary education more accountable to economic and job-related outcomes. Recent legislation such as 2012's "Student Right to Know before You Go Act," sponsored by Senators Wyden and Rubio, and the Obama administration's repeated commitment to a College Scorecard (Shear, 2014) echo a common demand for increased institutional disclosure of graduates' employment and earnings data in order to evaluate the public good of higher education. These join the 2013 college funding programs designed by the governors of Texas and Florida to increase access to state universities, discounted only for study in "workforce need" fields such as information

technology, logistics and management. Though these conversations take place in different arenas and highlight different concerns, they intersect profoundly in the question of the end and purpose of American higher education. Both conversations assert quite different central visions of which outcomes *matter most* in higher education.

For practitioners and supporters of liberal education, this point of intersection is of great concern. As liberal arts and humanities education becomes less sought after by students and their families (Lewin, 2013), this question is not simply an academic exercise but is instead a question of the very relevance of liberal education in the future of higher education. In order to remain relevant and viable in the American higher education scene, especially in the context of contentious public and political debate about the value of higher education, liberal education practitioners must give a robust and research-based account of the unique aims and outcomes they claim to offer.

Liberal education has long claimed moral education to be a chief mark of its educational format. Liberal education supporters regularly assert its unique ability to foster moral and ethical development in students, but this claim is insufficiently researched. Adherents of liberal education may attest to the great value of moral education for students and society, but data regarding higher education's efficacy in promoting moral development are limited. Additionally, the educational goal of moral development suffers important philosophical and epistemological critiques which bring into question its adequacy as a worthwhile aim in the contemporary era. In order to discern whether moral education is a suitable and sustainable objective for American higher education and hence whether higher education resources should be used to pursue it, supporters of moral education must identify clearly what moral education is, why it is

important for students, and how adequately it can be measured. This study seeks to contribute to that discussion through an examination of current scholarship in the field of moral development as it pertains to higher education and through a careful analysis of quantitative and qualitative data related to student development gathered in the evaluation process of a liberal education program. The central research questions of this project focus on aspects of student moral development and students' perceptions of the moral dimensions of coursework within a liberal education context. Certainly, no one study can articulate an entire educational program or pedagogical format, but the case for reaffirming moral education as an important aim of postsecondary education rests largely on a clear and persuasive defense of its impact in the lives of college students and its value in a democratic society.

1.2 MORAL EDUCATION'S HISTORY IN AMERICAN HIGHER EDUCATION

In early American higher education the cultivation of mind was undoubtedly tied to the cultivation of character and moral fortitude, and this fusion of intellectual and moral aims was valued for both its ecclesiastic and civic benefits. This construal of a moral dimension of higher education persisted even as the classical curriculum gave way to modern educational paradigms and philosophical shifts (Bowen, 2005; Brubacher & Rudy, 1997; DelBanco, 2012; Kiss & Euben, 2010; Reuben, 1996; Sloan, 1980). Educational historians note that the moral dimension of early American higher education depended on the centralizing notion of the coherent confluence of knowledge and the moral order (Brubacher & Rudy, 1997; Marsden, 1993; Reuben, 1996). In a recent piece on these historical foundations and their eventual breakdown, Reuben (2010) notes that the story of the decline of moral education as a centerpiece of American higher education

aims may be read as either of two distinctly different narratives: on the one hand, it may be seen as the attenuation of the religious and moral underpinnings of the foundational missions of American colleges, while on the other hand, the disintegration may be interpreted as the unshackling of higher education from the pervasive and invasive influence of religion and its moralizing capacity

From its beginnings American higher education was profoundly concerned with the moral character of its graduates. Colonial colleges served primarily to cultivate the minds and hearts of an elite class of gentlemen and a robust clergy for the New World and as such, held fast to their religious foundations and moral purposes. The capstone tradition of 18th and 19th century American colleges, inspired in large part by Scottish Enlightenment college models, “furnished an integrating principle for the entire curriculum” (Sloan, 1980, 5) that brought the college President into contact with college seniors to ensure that young men developed the sort of moral attitude and behaviors expected of educated, young Christian gentlemen of civil society (Brubacher & Rudy, 1997; Bowen, 2005; Kiss & Euben, 2010; Pascarella & Terenzini, 1991; Reuben, 1996). For the early founders of American higher education, the cultivation of mind was not simply an effort to address ecclesial needs but was grounded in a general interest in the cultivation of character and moral fortitude among the young men of this new society. Though moral or character education was part and parcel of the religiously driven missions of the colonial colleges it certainly had civic roots as well, as fewer than half of 17th century Harvard graduates went on to ministerial work (DelBanco, 2012). The aim of the early colleges’ cohesive classical and Christian curricula extended to whole person formation for the sake of a civic order populated with an adequate number of educated

men who could understand and judge wisely the world as given by God. However, the aim of character development and assessment became and for several centuries remained an admissions tool, a means of managing ethnic, gender and sectarian access to the early, elite colleges. This misuse of the notion of character as a means of exclusion may have proven to be part of its undoing, insofar as the rightful interruption of that abuse ultimately debased the role of character education in the academy.

As higher education expanded in the US in the 18th century, the diminishing power of churches over universities shifted the center of authority away from ecclesial bodies and increasingly toward autonomous faculties, college presidents, and eventually governing boards of colleges and universities. This shift ultimately dissembled the authoritative claim of a particular and highly focused aim of higher education (Reuben, 1996). Indeed, religious diversity within the early colonial colleges itself proved to be the first encroachment on various churches' authority over the aims and accounts of moral education, as internal divisions within Christian denominations ushered in discord regarding the elucidation of moral norms. By the end of the colonial period the pedagogy and educational content of the nine established American colleges remained under fairly strict church authority, despite the weakening effects of diversity on religious authority during this growth period.

It was in the post-revolutionary spirit that colleges were reconceived by the likes of Thomas Jefferson as a training ground for secular leaders, not simply clergy. In this new vision of a truly American form of higher education, curricular innovations and freedom from religious oversight and control were suggestive of a less centralized approach to the aims of higher education, offering Enlightenment virtues as the new

moral center. Already by the writing of his 1828 Yale Report, Jeremiah Day felt compelled to provide a vigorous apologetic of classical curricula's cultivation of "disciplines of mind" over "furniture of mind" in light of debates about the aims of education. Day admonished his colleagues to recall that the material advancement of a nation ought to be accompanied by an attendant moral advancement as well. The eventual dominance of 18th and 19th century science over and against natural theology contributed to higher education's skepticism about finding a robust grounding of morality. As empiricism and positivism prevailed in scientific inquiry, value-free research gained popularity among natural and social scientists. These shifts encouraged university reformers to endorse greater autonomy for faculty, departments, and students. The divestiture of university control of curricula stimulated great strides in academic freedom, curricular innovation and specialized scholarship but it also diminished the power of institutions to assert robust, university-wide educational goals. By the turn of the 20th century, journalistic exposés raising serious concerns about the moral environments of American colleges and universities prompted university leaders to take greater control of the various aspects of student life which had proliferated under the reform ethos of student autonomy and social freedom. By the 1920's, colleges and universities were substantially expanding professional staffs to deal with student life, shifting social and moral issues away from traditional academic contexts.

By the late 20th century the aim of moral and character education became particularly tenuous, largely out of sync with the aims of the research university model and besieged by a complex set of challenges from both consumers of higher education and the academy itself. The world wars of the first half of the 20th century precipitated a

persistent demand for scientific and technological advances and accompanying post-war anxieties about the social, political and ethical fabric of western democracy. The malaise following the World War I was only deepened by the recognition that the horrors of the Nazi regime flourished in a highly educated and civilized nation. By the start of American higher education's Golden Age (1945-1970), the emerging dominance of the research model and increasingly differentiated and highly specialized fields of inquiry brought into sharp focus the seeming irrelevance of moral education in a thoroughly modern university system. The Age of Sputnik ushered in a new, largely scientific agenda for universities attempting to answer the call of service to a common good situated firmly in the context of the Technological Age. Massive growth in higher education participation, funding, and collaboration with industry and government precipitated a major reconsideration of higher education's central purposes. Now, higher education was expected to produce knowledge rather than simply to transmit knowledge and culture from one generation to the next (Brubacher & Rudy, 1997; Kerr, 1963; Pascarella, Wolniak, Siefert, Cruce & Blauch, 2005; Reuben, 1996). Following the World Wars federal assistance programs introduced university aspirations to American families that had previously been excluded from the largely elite systems of the early collegiate model. Curricular changes in universities and colleges in the middle of the 20th century echoed the exigencies of the new research agenda of the technical age while the new egalitarian mode of higher education raised questions about the legitimacy of higher education's traditional aims and raised doubts that education *could* promote moral soundness in students. Powerful postmodern and feminist critiques of traditional agendas contested the overly rationalist and paternalistic nature of moral education and brought

into question its claims of ethical and moral normativity. Increasingly, student moral development was perceived as outside the purview of faculty or the primary work of a university, consigned regularly to general education electives, capstone courses, service learning courses and ethics modules within pre-professional coursework. By the 21st century even colleges that require philosophy and religious studies courses rarely offer courses with explicit goals of ethics or the advancement of students' moral character (Hoekema, 2010). Decades of research into the impact of college on college students in the second half of the 20th century and the beginning of the 21st have not quieted the restive contemporary conversations regarding the place of moral education in the American higher education scene.

1.2.A. Moral Education within Great Books Programs

In the midst of this 20th century unease regarding diminishing ethical norms, a new regard for a “general education” was developed in the Great Books movement. It sought to reintroduce moral education without reliance on sectarian traditions which were by then viewed as divisive to democratic society. In this context, University of Chicago's Robert Hutchins recast his Great Books movement, originally conceived as a rigorous intellectual program, as an educational model that offered the sort of character education needed for education of democratic citizens. Hutchins argued vociferously that the principles taught in Great Books programs provided students with robust understandings of democratic principles and laid the ground for moral as well as intellectual virtues and habits. Adopted by colleges like Harvard as an overview of and lively debate about historically important questions and texts (rather than Hutchins' vision of the program as formational of a moral grounding), the general education movement of the 20th century

became an influential educational venture. Along similar lines, Catholic colleges and universities largely retained a classical paradigm via the traditional Jesuit *Ratio Studiorum* curriculum model, despite pressure by accreditors to modernize and rejection of Catholic college-educated students in elite graduate schools. In the 1960's, programs designed to stimulate interest in moral and social values, such as experimental college curricula, residential programs, and seminar courses became a popular response to student movements' demands for curricular reform.

Student activism in the 60s successfully won greater student autonomy and student control over educational choices, largely subduing the momentum of the post-war general education movement. These movements challenged university *in loco parentis* practices and echoed larger societal shifts of increasing wariness of authority and a general hermeneutic of suspicion regarding moral norms. The secularization of many colleges and universities (and faculties) that were religiously founded or affiliated proceeded apace in this period. American higher education's great expansion during the 1960s precipitated rapid and significant institutional and structural changes that threatened its economic stability while public dissatisfaction with the civic mission of higher education continued to grow. As the economics of higher education became more precarious, student and administrative opinion regarding the aims of higher education tacked toward professional education and vocational training. Astin's long-standing survey administered by the Higher Education Research Institute of first year students' designation of what they hold most important sheds light on the shift (Reuben, 1996). By the late 80s, 80% of students considered being financially well-off very important or even essential, while only 45% considered "developing a meaningful philosophy of life" to be

very important, inverting the importance of those values from the late 60s. Finally, increasingly complex enrollment management systems developed in the last part of the century, coupled with rising reliance upon national rankings such as those found in *US News and World Reports*, fostered a consumer posture among students and their families. A university's ability to ensure students' professional and financial success replaced traditional goals such as moral and character education as primary pursuits of institutions. The dominance of economic concerns in American higher education has yielded only slightly to the aim of moral education in recent years, evinced by the rise and popularity of service learning programs, an increasing presence of ethics programs within professional schools, and the prevalence of "critical thinking skills" in the pantheon of higher education's central aims. American higher education's long and storied association with moral education could mean that recent attention to this facet of education might be another phase in a rather fickle relationship.

1.3 THE CONTEMPORARY DEBATE OVER MORAL EDUCATION IN THE ACADEMY

In recent years, growing number of leaders in higher education have strongly cautioned against American higher education's abjuring of liberal education and its moral development aims, calling for a reinvigoration of these traditional aspirations (Arum & Roksa, 2010; Bok, 1988, 2007; Delbanco, 2012; Hauerwas, 2010; Keohane, 2006; King, 1997, 2009; Kiss & Euben, 2010; Kronman, 2008; Lewis, 2007). Citing the needs of a flourishing and educated democratic citizenry, national higher education reports have likewise called for renewed attention to the aims of moral and character education. The Association of American Colleges and Universities report, *Great Expectations: A New Vision of Learning as a Nation Goes to College*, challenged higher education to foster

“an acceptance of responsibility for the ethical consequences of our actions and ideas” (2002, p. xii), reiterating the 1998 Higher Education Act’s affirmation of character development as one of the primary aims of American higher education. These exhortations echo former Harvard University President Derek Bok’s assertion of the obligation of educators to “help their students understand how to lead ethical, reflective, fulfilling lives,” in his continued advocacy of higher education’s role in fostering students’ moral growth (1988, 2007). Patricia King argues that educators ought to have a basic understanding of moral development and its processes, maintaining that the oft-cited mission statement goal of fostering good citizenship and character among college students requires the articulation and communication of “the moral dimension of university life” both in terms of the programmatic content and process (1997, p. 90). With a view to broader philosophical concerns, Martha Nussbaum recommends a revitalization of liberal education as an important corrective to the dominance of practical and economic influences in society, noting liberal education’s explicit aim of nurturing and developing the habits of mind necessary to relate to others beyond the confines of economic terms and relations (Gutmann, 1987, 1999; King, 2009; Nussbaum, 2004). *A la Dewey*, Nussbaum asserts liberal education’s unique ability to cultivate civic and narrative imaginations needed in deliberative democracy.

Not all educational leaders agree with these endorsements of the moral aims of higher education and liberal education’s unwavering commitment to it. One of the most ardent contemporary critics of higher education’s responsibility for moral formation is emeritus dean, professor of humanities and law, and literary critic Stanley Fish, who has been a passionate critic of moral purposiveness in the Academy. In his well-known

works, “Aim Low” (2003) and *Save the World on Your Own Time* (2008), Fish decries the goal of moral education as a faulty and unworkable idea that blurs the lines between the purely conceived academic purpose of higher education – to teach the content and techniques of specifically defined and discrete disciplines and train researchers for those fields – and dogmatic or partisan attempts to indoctrinate students to some predetermined notion of the good life. While he acknowledges that rigorous and intellectually grounded ethics courses are suitable for the academic realm, Fish maintains that this sort of critical assessment of various aspects of ethical theory or ethical dilemmas simply should not converge with aims of the advancement of the students’ moral lives. Fish points out that moral growth and development arises in and through so many variables as to make it impossible to evaluate, to plan for, or to teach to moral growth.

Here Fish makes an important distinction between ethics theory, which might be the proper purview of academic work, and ethical praxis. He suspects, rightly, that the moral education movement seeks to impact not only moral judgment but moral behavior in students, and this Fish rejects as a tenable aim of education. Central to Fish’s argument is the firmly held argument that higher education should resist tying its aims and goals to how a student turns out morally, ethically or civically. The moral growth of a student, Fish claims, is something for which no educator can or should be responsible; that type of development simply falls outside the bounds of education to facilitate, measure, or anticipate. Fish’s position is certainly not a new one. Plato’s *Meno* asks the same sort of questions, whether virtue can be taught and if so, by whom. Indeed, Plato concludes in the dialogue that virtue cannot be taught: centuries of “morally sound” education in the early years of American higher education cannot offer clear evidence that he was wrong

(Hoekema, 2010). Moreover, Fish's sentiments echo the seemingly unsettled quality of the moral education discourse among contemporary developmentalists and moral psychology researchers and are illustrative of a general attitude of many college and university personnel. Among even the most ardent supporters of moral development researchers, there is an acknowledged and problematic research gap between viable measures of moral judgment and moral action.

However, others within higher education leadership have recently proclaimed the need for a new form of moral education, one that eschews the traps of paternalism, parochialism, and ideology and rejects the dogmatic narrowness of Fish's position. Public intellectual and ethicist Stanley Hauerwas (2010) concedes that in the present postmodern context, in which we find little or no commonly held convictions about what is true or good, universities will necessarily struggle with questions of larger, less tangible purposes or educational goals because the terms and definitions of these are nearly impossible to identify. Hauerwas points out that the convergence of the specialization of disciplines and the professionalization of those disciplines has resulted in an extreme autonomy of academic fields which increasingly allows disciplinary justification only in and through the terms and relations of the field itself, thus diminishing a university's ability to define any central purpose or comprehensively construed moral aim. Interestingly, Hauerwas makes the claim that religious traditions, particularly in Christian and Jewish contexts, flout this narrowing and isolating principle since the emerging self-understanding of these traditions *as* ongoing and active, with public missions, demand a reflection on the dialogical roots of their central themes and debates. They also require, Hauerwas claims, an educated public to comprehend the various "agreements that make

their disagreements intelligible” (107). A similar case might be made for Great Books-style programs in liberal arts colleges and university programs, which also involve a uniquely dialogical and dialectical approach to understanding one’s own intellectual heritage and an attentiveness to one’s own cultural character. Hauerwas here picks up on a thread of an argument that has a long history in American higher education. Debates regarding the cultivation of an educated populace able to comprehend and appreciate the pillars of democratic society extend back to those of Socrates and the Stoics, Scottish Enlightenment thinkers and de Tocqueville. Indeed, the contemporary debates on moral education build on a long history of claims about what higher education can and should be expected to accomplish and what purpose it serves *vis à vis* the needs of the community.

Hauerwas’ and Fish’s arguments on the aims and purposes of higher education are much more than an academic exercise and are situated against the backdrop of criticisms lodged at higher education’s abnegation of moral education by educational leaders. Though conservative corners of the academic sector have long lamented the demise of moral education, the past two decades have seen criticism of this sort from across the political and institutional spectrum. Derek Bok, former president of Harvard, has been a vocal critic of higher education’s diminished commitment to the traditional connection between liberal arts and moral education since the 1990s. He has recently been joined in this protest by a cadre of equally respected and highly placed educators. The themes of Bok’s 2007 critique, *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More*, were echoed in that same year by *Excellence without a Soul: Does Liberal Education Have a Future?* by Harry Lewis,

former Harvard Dean, who published his own appraisal of the disjointed aims of higher education. Both Bok and Lewis decry the loss of educational cohesion in higher education. In Lewis' view, the decay of general education programs is deeply rooted in the entrenched specialization and isolation of departments and fields of inquiry and the subsequent deterioration of moral education as a goal recognized across disciplines. Other detractors offer similarly discouraging reviews of higher education, including former Yale Law School Dean Anthony Kronman's publication, *Education's End: Why Our Colleges and Universities Have Given Up on the Meaning of Life* (2008), sociologists and educational researchers Richard Arum and Josipa Roksa's *Academically Adrift: Limited Learning on College Campuses* (2010), and cultural critic and Columbia University Professor of Humanities Andrew Delbanco's *College: What it Was, Is, and Should Be* (2012). Numerous exposé articles in popular periodicals, including *Rolling Stone*, *The Daily Beast*, *The New Yorker* and *The Atlantic*, to name but a few, regale the public with tales of the moral dissolution of college and university students and the institutions that cultivate the student cultures in which they live and learn. Though these sorts of exposés seem perennial, they signal that American higher education is once again in a period of attention and concern regarding moral education's relevance.

1.3.A Contemporary Research and Moral Education

A retrieval of moral or character education as an aim of American higher education is a precarious endeavor. In many academic circles, this aspect of education is seen as parochial and paternalistic, while in others it is simply seen as impracticable and outside the jurisdiction of scholarly concern. Advocates of moral education recommend that higher education pay more than just lip service to the goal and criticize its nominal

use in university and college mission statements. A rapidly growing body of research, including several meta-analyses of recent research, confirms that moral growth and development is quite active during the college years and that collegiate experiences in particular impact moral growth significantly (King & Mayhew, 2002; McNeel, 1994; Pascarella & Terenzini, 2005). Indeed, Rest and Thoma's (1985) 6 year study of high school students, which included subjects who did not attend college and students who went on to two- and four-year colleges, found significant evidence of college's unique impact on moral reasoning development. Additionally, the large scale, 2006-09 Wabash National Study of Liberal Arts Education (WNS), culling data from 19 institutions of various types (liberal arts colleges, regional and research universities, and community colleges; private and public, single-sex and coed, religiously affiliated and non-religiously affiliated) found that among 12 outcomes measured, students' moral reasoning showed the largest positive gains. Indeed, these gains represented a ten percent increase in the first year, while many other outcomes showed little or negative change. This type of research raises questions about which aspects of the college experience precipitates moral growth and best practices for advancing and enhancing that development. Researchers in the field of moral development point out that the first year of college is particularly fruitful, finding evidence of what educational psychologist William Perry characterizes as a trajectory beginning in freshman year that moves a student from simplistic forms of ethical understandings, through relativistic configurations, to complex forms of commitments to values and meaningful horizons of living (Perry, 1970, 1981, 1999). As Patricia King and Mathew Mayhew assert within their own research in the field, moral reasoning is related to general cognitive structures

and does not simply unfold as part of the maturation process, context matters and the context of college matters significantly (2002).

In *How College Affects Students, Vol. 2* (2005), Pascarella and Terenzini revisit their iconic 1991 synthesis of over 2,600 studies done on the impact of college on college students, in which the net and long-term developmental effects of college participation were examined. This meta-analysis reconfirms significant change in college students' moral reasoning from freshman to senior year, findings which have been held across measurement instruments and different cultures and controlling even when controlling for subject maturation, socioeconomic status, and levels of precollege moral reasoning and intelligence. Pascarella and Terenzini acknowledge several impediments to any grand claims regarding growth in moral development, including the difficulty in assessing precisely the magnitude of effects, a lack of evidenced connection between moral reasoning and moral action, and research validity issues connected to the non-random assignating of college student subjects (since college attendance itself involves self-selection) along with regularly insufficient control groups. In his work on the intersection of moral development and higher education participation, researcher Mathew Mayhew similarly asserts that much work needs to be done to try to untangle the seemingly countless variables that converge to produce development of moral reasoning, but his work and his meta-analyses of research in this area reveals that progress is indeed being made in sorting out which aspects of college truly do account for the striking advances that college seems to promote in students' moral development (King & Mayhew, 2002; Mayhew, 2004a, 2004b; Mayhew & Deluca Fernandez, 2007; Mayhew, Wolniak &

Pascarella, 2008; Mayhew & King, 2008; Mayhew & Grunwald, 2008; Mayhew, Vanderlinden & Kim, 2010; Mayhew, Seifert & Pascarella, 2010, 2012).

Thus, three noteworthy trends converge related to moral education and its place in the academy. First, there is an expressed desire among educational and political leaders for more attention to be paid to moral education, and this desire is clearly connected to the goal of developing and strengthening the moral reasoning capacity of students. Second, research has shown clear evidence that the experience of going to college is uniquely influential on the advancement of moral reasoning and, though difficult, it is not impossible to identify which aspects of that unique experience contribute to this influence. Third, there is a decline in participation in and support of liberal arts or general education programs which often implicitly and explicitly aim to promote the sorts of intellectual skills and capacities needed for increasingly complex moral growth, such as critical thinking, the capacity for perspectival thinking, creative problem solving, exposure to ethical theory. Despite various reproaches regarding higher education's abandonment or embrace of moral development as an aim, moral education's place in higher education remains hampered by inadequate definition, content, and measurement. Recommending the reinvigoration of moral development as an educational goal necessitates a clear understanding of the various elements of moral development, as well as a research-based account of how moral growth might be measured and which aspects of higher education significantly influence that development.

1.4 THE AIMS OF THE PRESENT STUDY

The topic of this study regards questions about moral education: which aspects of student development it purports to impact, how to evaluate that impact, what place it

might hold in the pantheon of higher educational goals, and how to best assess various educational interventions' effectiveness in achieving those goals. Though there is little consensus about what constitutes moral education and what might count as outcomes or goals of moral education, great strides in 20th century developmental theory have enabled researchers to identify more clearly the scope of moral development and its many aspects. Educational research is well served by utilizing these advances to begin thinking more concretely about how to best promote moral development in students.

Though researchers concede that participation in college is a uniquely impactful experience on moral reasoning, it remains unclear how and to what extent particular educational interventions such as course content might influence the development and enlargement of moral reasoning abilities. This study examined the impact of an academic program's course content on the moral reasoning of students and the association of students' perceptions of the moral dimensions of course content and their own development. The course at the center of this study is an interdisciplinary, liberal arts course that follows a great books model offered to first year students only at Boston College, a mid-sized research university. Part of a more comprehensive, interdisciplinary program of study housed in the university's philosophy department and included in the core curriculum, the course is designed to engage students and faculty in tracing the roots of major philosophical and theological inquiries through close readings of primary texts. Heavily influenced by Heideggerian insights into historical-critical methodology, the course seeks to make students more consciously and critically aware of their own notions of what is good, true, and valuable for individuals and for communities. Taking up ethical theory explicitly at points throughout the course, students engage in an evolving dialogue

between ancient, medieval, modern, and contemporary thinkers about the moral aspects of personal, social and political orders. Textually centered discussions seek to encourage students to think critically and foundationally about the private and public nature of ethical and moral choices. Though the explicit aims of the Perspectives Program are primarily intellectual, the first year course, *Perspectives in Western Culture*, seeks to foster a critical and reflective consideration of the ways that a student's values have been shaped by her culture and history. Taking up the question, What Is the Best Way to Live?, as a central theme signals to both students and faculty that the course intends not simply an academic study of these texts but also aims to establish and invite students into a robust conversation about how we understand, choose, and live our individual and communal values.

Whether or not liberal education courses like these are effective in advancing intellectual and moral development in students is crucial to the relevance of liberal education's aims in colleges and universities, particularly within institutions with research agendas and aspirations. Programs like the Perspectives Program must clearly articulate their education goals and find ways to demonstrate that they can deliver on the outcomes they pursue. Making grand claims rooted in education's nostalgic past or abstract notions of what we suspect and hope is happening in the hearts and minds of students is inadequate against the backdrop of rising costs of higher education and daunting economic uncertainties facing college graduates. If liberal education claims to make substantial contributions to democratic society and to the lives of young adults, it behooves practitioners to provide substantial research to showcase those contributions

and to identify best practices that make those contributions possible. The present study seeks to aid in that endeavor.

Chapter 2:

Literature Review

Once a central aim in the lecture halls and campus cultures of American higher education, moral or character education has been largely jettisoned from college classrooms and residence halls and relegated to university mission statements. Recent conversations among politicians and leaders in higher education, however, have recommenced a public debate on the fundamental aims of American higher education, opening the door to a renewed consideration of this once preeminent educational aspiration. Practitioners of liberal arts education have long claimed moral education to be a chief mark of its educational format and assert its unique ability to foster moral development in students, but this claim is insufficiently researched. While supporters of liberal arts education attest to the great value of moral education for students and society, data regarding higher education's efficacy in promoting moral development are limited. As a national discussion moves forward about what matters in higher education, it behooves those invested in liberal arts education to demonstrate through empirical research that moral education affects moral and ethical development, that this growth can be measured as such, and that aspects of liberal arts education significantly influence moral development.

Advances in the field of moral development in the last half of the 20th century provide a foundation for contemporary work toward these goals. Piaget's and Kohlberg's early theoretical structures have given important stability to the field, allowing for burgeoning contemporary research into various aspects of moral and ethical

development. Neo-Kohlbergian studies and research are beginning to lay the groundwork for a thorough exploration of aspects of moral development such as ethical sensitivity, moral reasoning and principled action, as well as their connections to education. This literature review seeks to examine how a key piece of moral development—moral reasoning—has come to be understood relative to moral development generally, how adequate measurement tools have been designed to assess moral reasoning, and how research may help educators better understand and promote moral development in their students. What follows is a review of the literature of these topics to gain an overview of the present state of the field.

2.1 MORAL DEVELOPMENT RESEARCH

In its 2002 report, *Great Expectations: A New Vision of Learning as a Nation Goes to College*, the Association of American Colleges and Universities challenged higher education to foster “an acceptance of responsibility for the ethical consequences of our actions and ideas” (p. xii), echoing former Harvard University President Derek Bok’s 1988 assertion of the obligation of universities to “help their students understand how to lead ethical, reflective, fulfilling lives,” in his piece on higher education’s role in fostering students’ moral growth. To these ambitious and yet ambiguous educational goals, Patricia King adds that educators should have a basic understanding of moral development and its processes, noting the helpful neo-Kohlbergian conceptual frameworks in mapping out common patterns of moral and ethical growth (1997). King argues that the goal of fostering good citizenship and character among college students, an aim regularly highlighted in the mission statements of institutions of higher education,

requires the articulation and communication of “the moral dimension of university life” both in terms of the programmatic content and process (1997, p.90).

Working from Kolbergian and Piagetan foundations, neo-Kohlbergian researchers James Rest and his colleagues at the University of Minnesota (referred to henceforth as the Minnesota group) have contributed more than thirty years of research on the moral development of college students and young adults. Rest’s interest in the psychological and cognitive aspects of moral behavior led the Minnesota group to examine how the various components of moral development evolve and function. Rest’s work provides three important tools for the advancement of moral development research: a schematic of moral development that attempts to revise Kohlbergian and orthodox stage model paradigms by employing softer “schemes” or patterns of moral reasoning; a four-component conceptual model of moral development which allows researchers to focus more precisely on specific functions of moral development; and a much tested and widely accepted measure of moral judgment, known as the DIT (Defining Issues Test, currently being used in its second version, the DIT2, hereafter referred to simply as the DIT). The Minnesota group’s work has advanced moral development research significantly and the DIT’s efficacy, facility, and large norming samples make it a highly reliable and much used measurement device. The DIT aims at capturing patterns or modes of moral reasoning, identified by Rest as one of the most salient features of the much more broad set of functions properly understood as moral development. Rest’s central question regards how moral awareness ignites motivation and moves toward action and character via moral reasoning, decision-making, judgment and perseverance. In his early work, Rest points out that moral development is itself not necessarily a predictor of moral

action, but rather, it is “a necessary but not sufficient condition for moral action” (1986, p.58). Thus, he began to conceptualize moral development not as a unitary process but a dynamic set of interconnected components, each of which might be observed or assessed in patterns of usage and growth. Rest postulated that Kohlbergian attempts to measure moral development via the Moral Judgment Interview, developed by Kohlberg during his doctoral work in Psychology in 1958, failed to distinguish adequately the non-sequential, non-linear fluctuations of the various processes of moral reasoning.

Kohlberg’s work relied heavily on the foundational work of Jean Piaget and was thus influenced by Piaget’s groundbreaking work in exploring and explaining the cognitive developmental processes of children as they navigate moral meanings and judgments (Piaget, 1997). Piaget pioneered a framework for developing a theoretical construct of moral development, identifying the basic cognitive and logical structures that children use to sort out moral meanings and construct moral decisions and values. Piaget’s work departed from a Durkheimian “Character Education” or cultural socialization model, which posited that students develop moral reasoning through didactic methods and positive reinforcement of good behavior (Snarey & Samuelson, 2008). Piaget’s model insisted on a view of the child as an active participant in her own moral thought and action via cognitive developmental processes that emerge out of action and interaction with one’s culture and environment. Piaget observed that children meet increasingly complex moral situations that challenge the limitations of their own structures of thought and understanding, and postulated that these “collisions” would instigate a demand for higher order thinking. Construction of these higher orders, Piaget

asserted, was the central moral task of a child, and he argued that these constructions were cognitive and structural in nature.

Piaget's approach also diverged from the behaviorist tradition's focus on causal correlations of human action and behavior, attending instead to verbalizations and explanations not as behavior indicators or predictors, but as the expression of a world of meaning and value out of which behaviors flow (Rest, 1979). For Piaget, verbal and written responses of children revealed not simply behavior motivation, but more importantly illuminated a developing inner world of children, with a particular internal logic and sets of meanings and values. Piaget did not assume a world as "given," with human understanding striving to work out better and more direct apprehension of that world, but rather he attempted to sort out the conceptual framework and subsequent intuitive position of a child that *is* "the world" of the child's moral point of view. For instance, Piaget proposes that a child's behavior reflects his or her own understanding of moral obligations as fixed frameworks, characterized by Piaget as "moral realism," seen not as social constructions but as fixed laws. The movement from heteronomous morality, in which a subject unilaterally conforms to authorities and rules, to an autonomous morality that includes reciprocity and mutual cooperation, demonstrates for Piaget an important evolving cognitive development. This development undergirds a conceptual framework needed to scaffold increasingly complex moral problems and their solutions. Using stories and games, Piaget elicited and observed children's verbal accounts of behavior, noting unique elements from which he drew conclusions about underlying cognitive and sense making operations. Piaget's work in identifying age-related differences capitalized on advancing complexity and nuance in cognitive

operations. But Piaget postulated that development of reasoning capabilities ended after adolescence, and so it was left to later researchers like Lawrence Kohlberg to explore the further development of moral reasoning in post-adolescence.

2.2 LAWRENCE KOHLBERG'S WORK AND INFLUENCE

Kohlberg's work with young adults, primarily college students, found its roots in his own pre-college experiences during and after the Second World War. Having learned of the atrocities of the Holocaust in high school, Lawrence Kohlberg postponed his college career, traveling to Europe to witness the end of the war (Snarey & Samuelson, 2008). His Zionist sympathies stirred, the young Kohlberg engaged in illegally smuggling Jewish refugees through a British blockade, likely rousing a formative set of moral questions via a unique and moving experience of a personal moral dilemma. Quickly completing his undergraduate degree at the University of Chicago upon his return to the US, Kohlberg moved on to a doctoral program in psychology where his dissertation work centered on a single moral dilemma and the responses elicited by it in a group of adolescent boys. Noting significant age-related differences, Kohlberg posited a six stage theory of moral judgment which drew heavily on the type of cognitive developmental work put forth by Piaget (Reed, 2008; Whitely, Bertin, & Berry, 1980). In later years, Kohlberg would also retrieve and utilize aspects of Durkheim's character education, or cultural socialization model, adverting to the need for a democratically formed version of indoctrination. According to this model, in the context of shared and respected rights education by a social collective would not seek to extinguish the priority of the individual and would therefore allow an individual's autonomy in the face of an illegitimate authority. In establishing a proper tension between the participant-centered cognitive

developmental model of Piaget and the authority granting cultural socialization model of Durkheim, Kohlberg was convinced he had sidestepped the basic problem of extreme cultural relativism the likes of which might occur when a highly educated, sophisticated social community surrenders its moral reasoning capacity to a totalitarian regime.

Kohlberg's research involved the design and implementation of the Moral Judgment Interview (MJII) in which college-aged subjects were asked standardized questions regarding a hypothetical moral dilemma. The interviews elicited subjects' explanations and verbalizations of judgments in attempting to resolve the dilemma (Colby, 1983; Colby, Kohlberg, Gibbs & Lieberman, 1983; Kohlberg, 1981, 1984; Pacscarella & Terenzini, 1991; Rest, 1999). The subject of much criticism and acclaim, the MJII nevertheless generated a wide array of studies on the impacts of age, gender, race, ethnicity, etc. on moral judgment, and contributed substantial research on the impact of college on students' moral development (Brabeck, 1984; Gilligan, 1982; Nucci & Pascarella, 1987). Using standardized questions and classification methods of analysis of interview responses, Kohlberg asserted that he could observe and identify direct indications of the cognitive processes involved in a mode of moral reasoning and could thus determine a "score" that identifies a stage of moral development within a continuous scale (Pascarella & Terenzini, 1991; Rest, 1999b). Kohlberg's work with the MJII was grounded in his basic allegiance to a social-construction theory – an assertion that the self emerges in and through series of patterned responses to social interactions – and that this self-constitutive process is mediated by cognitional structures that are reconstructed as interactions between the self and others become more complicated and more demanding (King, 2009; Reed, 2008). In light of this social constructivist underpinning, Kohlberg

asserted that direct and repeated social-moral experience was indeed necessary for moral development, motivating him to advocate for educational interventions like role-playing exercises and moral dilemma discussions (Armon & Dawson, 1997; Kohlberg, 1984; Reed, 2008). Via research data from the MJJ, Kohlberg hypothesized that subjects move through basic stages of moral development in a structured, invariant pattern. Like Piaget, he noted that progression through the stages was fueled by the subject's recognition of the inadequacy of existing structures of thought and judgment. Facing complex challenges, he observed, would precipitate this recognition and the subsequent need for higher orders of thinking in predictable and identifiable patterns. These patterns provided Kohlberg with the grounding of his renowned six stage developmental model of moral reasoning, progressing from a morality centered on the self to an other-centered morality, which can recognize and accept the perspective of another person or group of persons, known or even unknown to the subject.

For Kohlberg (1984), a subject's moral development may be traced through three levels which include six "hard" – in the sense that they are discrete and sequential – stages of cognitive patterning, each of which might be characterized by a particular moral orientation. In the egocentric "pre-conventional level," we find the Punishment-Obedience stage (Stage 1) and the Instrumental-Relativism stage (Stage 2) in which rewards and punishment, rule following, fairness as equality and reciprocity, and self-interest dominate moral judgment. Moving sequentially to the multi-perspectival framework of the "conventional level," a subject may advance to Kohlberg's Interpersonal Concordance stage (Stage 3) and the Law and Order stage (Stage 4), in which relationships and mutuality temper self-interest, allowing for consideration of a

third person perspective, adverting to the significance of duty and obligation, and accepting the social norming that supports these kinds of social and political arrangements. Kohlberg's final "postconventional level" posits a Social Contract/Legalistic Orientation stage (Stage 5) and a Universal Ethical/Principle Orientation stage (Stage 6), in which complex and nuanced understandings of justice and morality give rise to apprehension of universal sets of rights and principles. Kohlberg notes that these stages track the cognitive-structural path of moral development that is not separate from but rather parallels affective aspects of development, supporting other researchers' suggestions that moral development is best understood as one part of an integrated network of development (Bruess & Pearson, 2000; Pascarella & Terenzini, 1991). Initial use of Kohlberg's MJI showed promise with college aged students, though meta-analysis of data from early versions of the MJI give very limited evidence of third level (Stage 5 and 6) reasoning and startlingly little increase in moral reasoning in older adults, raising questions about the efficacy of the measurement tool and the adequacy of the stage theory itself (Armon & Dawson, 1997; Whitely, Bertin, & Berry, 1980).

Kohlberg's basic theoretical construct has many supporters and detractors, including claims of male bias in Kohlberg's evaluative model (Gilligan, 1982) as well as insufficient evidence of stage 5 and stage 6 reasoning (Rest, 1999b). The MJI as a measurement tool also suffers from demonstrated limitations of interrater reliability of interviewers and interview techniques and self-reporting biases (Kay, 1982; Nisbett & Wilson, 1977). Additionally, later researchers including James Rest cite fundamental problems with the "production" model of the MJI, noting that its demands on subjects' verbal judgments skew assessment of cognitive structures and sensitivities. For these

reasons, James Rest and his colleagues (1975, 1979, 1981, 1986, 1999, 1999b) sought a more reliable and effective measurement tool which would build on Kohlberg's basic moral dilemma structure and sidestep the limitations of the MJT's production mode and interview analysis format while also attempting to address the limitations levied against Kohlberg by other critics.

2.3 NEO-KOHLBERGIAN RESEARCH: JAMES REST AND THE MINNESOTA GROUP

Like Kohlberg, James Rest's own personal history precipitated a search for explanations of moral and ethical dissonance in a tumultuous moment in American history. It also led him to assess the limitations of Kohlberg's work measuring moral development. Raised by a minister in the American South during the Civil Rights Era, Rest was dismayed at the inability of members of the family's church congregation to respond to the legitimate claims of the civil rights movement in a way congruent with the general moral behavior of that congregation (Thoma, 2002). Rest wrestled with the discordant attitudes of people he knew to be otherwise charitable and generous, leading him to conclude that moral engagement might be more broad-ranging and situation-dependent than Kohlberg had envisioned. Rejecting the hard stage quality of Kohlberg's theory, Rest instead opted in his own line of inquiry for "softer" schemes, seeking patterns of interconnected components that together make up the broad spectrum of moral development. Rest joined others (Bebeau, 2002; Rogers, 2002; Thoma, 2002) in noting the broad range of what is commonly held to be morality and moral development, choosing to focus on a measure of moral judgment, considered by Rest to be the most pivotal component of moral reasoning and moral behavior (Rest, 1986).

By the early 1970's, Rest had left Harvard for a position at the University of Minnesota. At the time, Kohlberg and his colleagues at Harvard were beginning to turn their attention to the difficulties of the interview and scoring methods of the MJJ, eventually jettisoning scoring for Stage 6 due to lack of data. Rest sensed that the interview method, held onto by the Harvard group in the face of growing criticism, depended too heavily on a production model of justification from subjects. At the University of Minnesota, Rest reworked the basic format of the MJJ, developing a paper and pencil test in which subjects read and responded to a series of dilemmas, including but not limited to the Heinz dilemma of the MJJ. Using research from MJJ studies, Rest noted that at certain stages, subjects noticed and utilized particular elements of a dilemma in their attempts to resolve the conflicts (Rest, 1986; Thoma, 2002). This moved Rest and a growing contingent of colleagues in Minnesota to develop prototypic statements (with accompanying irrelevant non-stage typed statements) to prompt subjects to respond to various stage-related justifications for various resolutions. Thus, unlike the MJJ model, the DIT asks subjects not to *produce* justifications for their attempts to resolve conflicts and sort out the complexities of dilemmas, but to choose among statements that present various angles on the dilemma. The six dilemmas in the DIT are accompanied by a set of twelve statements that a subject might take into consideration in thinking about a possible resolution of a moral conflict. The subject is then asked to rate on a five point Likert scale the degree of relative importance of each statement in the subject's consideration and decision regarding the dilemma. Finally, the subject is asked to rank the four most important and motivating statements from the list, providing a "second pass" through the statements and giving a second mode of scoring the subject's responses. In this way, the

DIT does not rely on the *production* or *articulation* of moral reasoning, which in the MJJ risked confounding verbal ability and advanced moral development, but depends instead on the *recognition* of aspects of moral reasoning pertinent to working out moral issues. A series of scores are gathered from the rating and ranking tasks to produce a Principled Reasoning score (P-score), indicating the extent to which a subject used principled reasoning (later used to denote postconventional) in working out the dilemmas. This first version of the DIT was published by Rest in his 1979 work, *Development in Judging Moral Issues*. This early work developing the DIT would lead James Rest and the Minnesota group in a number of different research directions, three of which are significant for research in moral development and for the present study.

2.4 THREE NEO-KOHLBERGIAN RESEARCH DIRECTIONS

First, the DIT proved to be an excellent research tool for measuring moral development or some aspect of it. It was easy to administer and to score and thus became a regularly used tool, which in turn enhanced the Minnesota group's ability to demonstrate its effectiveness and validity. Over the years, revisions of the DIT itself and scoring techniques associated with the test (Rest et al., 1986, 1999) have resulted in remarkable advancements in moral reasoning research and have fortified the stability of the DIT as a reliable research tool. Second, it gave the Minnesota group a set of data with which to reassess Kohlbergian stage theory, eventually leading the group to work with schema theorists to reconceive Kohlberg's "hard" stages. What emerged in the 1980s for the group was a "neo-Kohlbergian" moral schema theory, diverging from an "orthodox" stage model while retaining some salient features of Kohlberg's fundamental theory. Third, using emerging data from the DIT, the group began to address strong criticisms of

moral development theory's inability to bridge the apparent gap between moral judgment and moral action, levied ardently by critical reviews of moral development literature (Blasi, 1980; Schlaefli, Rest, & Thoma, 1985; Thoma, 2002). In response to the assertion that moral action must necessarily serve somehow as the final arbiter or measure of moral development, the Minnesota group began to work on a formulation of a more robust and precise conceptualization of moral functions and their relevant domains within development generally. The result of this work was the Four Component Model, outlined by Rest in 1983 in a review of literature of moral processes for a series volume on cognitive development, in which Rest articulated findings of differentiated domains of moral competency and adverted to affective as well as cognitive developmental demands of moral maturation (Rest, 1983; Thoma, 2002). Each of these three research directions has proved extremely fruitful for late 20th century and early 21st century research in the area of moral development generally and moral reasoning specifically.

2.4.A. Development of the DIT

Rest and his colleagues have contributed more than 40 years' worth of expansive research with and on the DIT and its extensive use by researchers in widely divergent settings with a variety of populations has added to its reputation as a robust research tool (King & Mayhew, 2002; Rest, Thoma, Narvaez, & Bebeau, 1999). Rest's exposure to developments in contemporary Psychology convinced him of the limitations of hard stage theory for the test's evaluator purposes and eventually led him to take up insights from schema theory in developing his own "soft stage" model (Thoma, 2002). The Minnesota group retained Kohlberg's foundational assertion that growth in moral judgment is a cognitive development, and hence anticipated observable upward movement in moral

reasoning that would bear out with age, would map on to other age-related development, and would be particularly correlated to intellectual development. Further, in agreement with Kohlberg the group asserted that specific experiences, particularly socio-moral experiences that demand complex modes of thinking, ought to influence the progression of complex moral reasoning.

The search for robust indices able to capture the sort of moral reasoning growth that the group identified was a grounding task in the early years of the DIT formulation. P scores were initially used to denote a subject's "principled considerations" ranking based on Kohlbergian post-conventional prototypic responses (Rest, Thoma, & Edwards, 1997). The DIT's use of *recognition* formats over the MJJ's *production* format resolved the conflation of cognitive and moral development and allowed for tacit and unarticulated principled reasoning to be assessed.

Attending to limitations acknowledged by the researchers themselves as well as criticism from other researchers including Kohlberg (1979), the Minnesota group sought an index that would give better results without having to give up decades' worth of DIT data. By the late 1990's the group, working with over two decades of DIT-based research, gleaned insights into the relative power of the P score, rejecting others that failed to outperform it (Davison, 1977; Evens, 1995; Lawrence, 1987; Thoma, 1994, 2002; Thoma, Rest & Davison, 1991). Eventually, Rest was able to detect a more precise index, known as the N2, which combined P scores with decreases or systematic rejections of lower stage reasoning (Lind, Hartmann & Wakenhut, 1985; Rest, Thoma, Narvaez, & Bebeau, 1997). This combined power of two effects allowed researchers to identify when subjects are simultaneously gaining complexity of moral reasoning and clarity in

rejecting overly simplistic or inadequate reasoning. This new score has offered a robust mode of observing the influence of educational interventions on the moral reasoning capacity of students (Bebeau & Thoma, 1994; McNeel, 1994).

King and Mayhew's extensive meta-analysis of DIT research (2002) reviewed over 500 publications (peer reviewed articles, conference papers, dissertations, etc.) that have used the DIT in research with college students, though they noted that a substantial number of these studies used college students merely as a proxy for intelligent young people and many of the studies did not primarily use the DIT for an examination of moral reasoning. Focusing on 172 of these studies which used the DIT specifically to explore the impact of undergraduate college experiences on moral development, King and Mayhew's analysis strongly supports Rest's assertion of the DIT as a robust measure of moral reasoning and his claim of its responsivity to the impact of educational interventions.

2.4.B. Schema Theory and the DIT Schemas

The new scoring techniques of the DIT represented an important shift in the Neo-Kohlbergian work of the Minnesota group, which in the late 1990s began utilizing schema theory rather than Kohlberg's hard stage depictions. The N2 score attempts to identify "shifting distributions of stages" as opposed to hard or discrete stages, thus locating the extent to which a subject *tends to use* higher or more complex levels of moral reasoning (Rest, Thoma, Narvaez, & Bebeau, 1997; Thoma, 2002). In this way, the Minnesota group sought via the N2 scores not to pinpoint a stage of moral development, but to "assess the pattern of responses across stage orientations, estimating development on a low to high continuous scale" (Mealy, 20xx, p. 40). This use of softer patterns along

a continuum depended heavily on Rest's interest in emerging work in Psychology in schema theory.

Rejecting Kohlberg's limiting hard stages, Rest and his colleagues turned to Schema Theory which is "concerned with the application of organized generic prior knowledge to the understanding of new knowledge" (Rest, Narvaez, Bebeau, Thoma, 1999). Conceived by theorists as cognitive structures built throughout an individual's development through experiences and her reactions to them, schemas provide a conceptual framework used by an individual when confronting new data or questions. These frameworks serve to parse out the various aspects of the new data, attempt to fill in missing information, and guide the individual in the pursuit of further relevant information toward a solution or goal with respect to the data. Rest found Schema Theory more helpful than the operations-based stage theory employed by Kohlberg, which even Kohlberg lamented for its restrictions (Kohlberg, 1984). Though Rest himself worried about the DIT's abstract use of schema theory and wondered if the term *schema* might prove inadequate in identifying what the DIT captures (1999), he nonetheless saw the construction of the DIT's moral dilemmas as a tool to activate moral schemas in such a way that researchers could evaluate an individual's *working conceptions* of basic moral principles.

Recent developments in cognitive science ratify the Minnesota group's choice of schema theory. Accounting for the ways that individuals organize experiences and move toward interpretation and articulation of those experiences, cognitive scientists identify conceptual structures that utilize experiential memory and past understandings to receive and flexibly handle present experiences. Schema theory suggests that "general knowledge

structures residing in long-term memory” are formed by patterns or experiences layered with interpretation and meaning which in turn form conceptual structures that anticipate, receive and attempt to interpret new experiences and data (Narvaez & Bock, 2002, p.300). Likewise, memory functions in such a way to create cognitive fields that correlate various experiences into working categories which might receive new data, allowing a variety of relevant understandings and assumptions to come to bear on decision making as new situations arise for an individual’s consideration (Derry, 1996; Narvaez & Bock, 2002). Schema theorists posit overarching structures, or “mental models” that integrate these various memory caches and their correlated cognitive fields into meaningful and explainable horizons of decision and action. Rest’s use of these various aspects of schema theory was considerable (Rest, 1979; 1999b). He noted two crucial attributes of schemas, namely, that schema progress tends to be flexible and dynamic (as opposed to stage progress), and that while backward and forward progress through schemas is common, as individuals activate later schemas they are less likely to utilize early schemas. Moreover, he noted that schemas may be activated without an accompanying ability to articulate or explain the reasoning behind a chosen path of solving a dilemma. Rest concluded that the DIT and its verbal recognition model would thus resolve the weaknesses of the MJJ’s verbal production model and its hard stage restrictedness.

Narvaez and Bock point out that Rest’s early dissertation work adverted to three tasks which together form the movement of development and would later be the basis of his shift to a recognition measure: preference, comprehension, and spontaneous production (Narvaez & Bock, 2002; Rest, 1973). By presenting subjects with prototypic statements which reflect the moral reasoning of the various Kohlbergian stages, Rest

could elicit from subjects their evaluation of and preference for each stage. What Rest began to understand was that moral development moves in ordered, hierarchical patterns and that movement toward a higher stage begins with increasing *preference* for the higher stage reasoning, which might be identified in rating and ranking tasks with fragments of stage reasoning. This preference is followed by comprehension and the subsequent ability to produce justification or articulation of higher stage thinking. In the end, Rest and his colleagues chose the recognition model with accompanying rating and ranking tasks exclusively, since production models confounded moral development with verbal ability and comprehension skills.

The DIT focuses primarily on macro-moral questions, highlighting how social and institutional orders operate with particular emphasis on decisions and actions regarding those not in our own social or personal spheres. The schemas conceived by the Minnesota group (1999) were closely related to but not tidily mapped onto Kohlberg's original 6 stages, which Kohlberg grouped evenly into three levels (Pre-Conventional, Conventional and Post-Conventional), identifiable via cognitive operations. The DIT's extensive use of verbal recognition patterning via a set of read moral dilemmas excluded most of the earlier childhood, stage 1 individuals. The remaining early stages are collapsed by Rest into a schema known as the *Personal Interest Schema*, which closely parallels Kohlberg's Stages 2 and 3 (see Table 2.1), and highlights cognitive fields of personal interest and advantage. The collaboration and cooperation of this schema give way to give way to reciprocity and care for others. The *Maintaining Norms Schema* emerges via an interplay of understandings of cooperation, fairness, and reciprocity that intersect with conceptions of those outside the "in-group" (Narvaez & Bock, 2002; Rest

et al., 1999). Shifting to this schema relieves the tension of situations and questions insufficiently answered in the Personal Interest Schema, finally addressing the exigencies of the demands of “the Other,” those outside the scope of the group known to us. Kohlberg’s stage 4 falls within the range of this schema, which involves the cognitive coordination of larger social demands and meets the need for functional societal systems of rules, codes, and laws. Finally, Rest’s *Postconventional Schema* envelops stages 5 and 6 of Kohlberg’s theory, stages which partially or wholly eluded measure via the MJJ. This final schema encompasses individuals who, while committed to the primacy of shared and shareable ideals, advocate these ideals while adverting to contextual exigencies of communities, cultures and times. Rest is less precise in defining this schema than Kohlberg but includes in it characteristics of full reciprocity and moral purposiveness (as opposed to *de facto* norms) as necessary components in moral judgment at this level.

Table 2.1 Kohlberg’s Stages and Rest’s Schemas

Kohlberg’s Stages	Rest’s Schemas
1. Punishment-Obedience stage	NA
2. Instrumental-Relativism stage	<i>Personal Interest Schema</i>
3. Interpersonal Concordance stage	
4. Law and Order stage	<i>Maintaining Norms Schema</i>
5. Social Contract/ Legalistic Orientation stage	<i>Postconventional Schema</i>
6. Universal Ethical/Principle Orientation stage	

2.4.C. The Four Component Model

Noting the accepted view of Pascarella and Terenzini on college's general impact on student development (1991, 2005), King and Mayhew (2002) point out that much work needs to be done in defining and delimiting the specific domain of moral development. Rest's work in developing the Four Component model of morality enabled the Minnesota group to hone in on specific facets of moral development that uniquely incorporate cognitive and identity development (King, 2009; King & Mayhew, 2008). Recently updated by Bebeau and Monson for research in professional education (2008), the Four Component model makes clearer which aspects from the larger domain of morality might be addressed and observed in research studies. Diverging in large part from previous models of morality which tended to separate the cognitive, affective, and behavioral aspects of morality, Rest and the Minnesota group posited four components of morality each involving interactions of cognition, affect and behavioral manifestations (Walker, 2010). The model addresses the various psychological aspects of moral functioning, and allows research to target specific areas of moral growth or deficiency. The Four Component Model adverts to the "multiplicity of processes" involved in the psychology of morality (Rest, 1999b, p. 100) and identifies four components that together make up the inner psychological landscape of morality: moral sensitivity, which involves the ability to recognize the moral dimension of a situation and the capacity to see the impact of a situation from another's point of view; moral judgment or reasoning, a subject's ability to assess the implications of decisions and actions and an accompanying understanding (even if only notional) of the underlying criteria of moral choices; moral motivation, denoting the level of commitment one has to a set of chosen values and the

courses of action they call for; and finally, moral character, the degree to which a subject persists in implementing and following through on moral decisions and tasks (King, 2009; Rest, Narvaez, Bebeau, & Thoma, 1983, 1999).

Via this distillation of the various aspects of moral development, the Minnesota group pointed out that Kohlberg's work deals primarily with the second of these components, moral reasoning, and suggest that a subsequent, neo-Kohlbergian approach might continue to explore that element of moral development by bringing new research in cognitive and identity development into dialogue with Kohlberg's groundbreaking work. The Four Component Model also blunted a debate among researchers about the apparent gap between moral reasoning and moral action that had arisen in the 1980s (Blasi, 1980). In distinguishing the various aspects of moral processes, Rest and his colleagues were able to sharpen their range of research questions and more clearly identify the significance of DIT findings. Similarly, King (2009) notes that this demarcation of the proper purview of Kohlbergian research helps to address and possibly counter many of the critiques of Kohlberg's work, noting for instance Brabeck's claim (1983) that the ethics of care dispute raised by Gilligan is partially resolved in the component model, since care and empathy issues are more adequately understood as part of the first component, moral sensitivity. The component model also offers helpful distinctions for present contentious public discourse regarding education's seeming failure to promote sound moral reasoning and engender moral development, when, as Christian Smith points out in his study of emerging adults, young adults seem to be unable to determine in any meaningful way what it is that makes a moral issue *moral* (Smith, 2011).

Members of the Minnesota group have gone on to research the various specific components. Bebeau's work (1987, 1993) with professional school students resulted in the development of a measure of the moral sensitivity component (the Ethical Sensitivity Test) through which she has found evidence of significant variability among college students within this component. Her work suggests that educational interventions such as ethical training programs may impact this component of moral development. Similarly, Narvaez's work (1998, 1999, 2001) contributes much in drawing out salient features of moral comprehension on various developmental aspects of morality, such as the impact of personal and cultural background, the modes and movements of tacit knowledge, and the interaction between moral development levels and sensitivity. Walker notes, however, that while advocates of the component model advert to the cognitional and affective aspects of the components, most research rather narrowly highlights the cognitional, leaving a gap of insights into the affective pieces of the various components (2010).

2.5. A NEO-KOHLBERGIAN DEFINITION OF MORALITY

The theoretical lens used in this study includes a neo-Kohlbergian understanding of morality that James Rest and his colleagues posited in the last decade of the 20th century and has been used in neo-Kohlbergian research ever since. Rest asserts that moral reasoning is a "psychological construct that characterizes the process by which people determine that one course of action in a particular situation is morally right and another course of action is wrong" (Rest, Thoma & Edwards, 1997). While he affirmed that this capacity includes a cognitive capacity, but also suggested that the cognitive strategies we use vary significantly from one stage of development to another (Rest, Thoma & Edwards, 1997). It is this variation that Rest and more recent neo-Kohlbergian research

have sought to investigate in order to apprehend more clearly what capacities are used (or not used) as we approach increasingly complex and nuanced moral and ethical issues.

Rest surmised that when the Maintaining Norms schema is preferred, subjects will display great preference for giving unlimited power to authorities at the expense of individual rights or needs and will prefer clear and possibly even simplistic social norming practice, whereas those with Postconventional schema orientations will favor the needs and rights of individuals despite their acknowledgement for systems, norms and public policies that function well. They will demonstrate preference for systems and norms that allow for individual rights to be expressed and addressed and that serve the collective will rather than make unnecessary demands of it (Rest, Narvaez, Bebeau & Thoma, 1999). Thus, the neo-Kohlbergian notion of Postconventional thinking, or the highest order thinking concludes “that rights and duties are based on sharable ideas for organizing cooperation in society, and are open to debate and tests of logical consistency, experience of the community and coherence with accepted practice (Rest et al., 1999, p. 41). This position stands apart from a number of other streams of moral theory, including deontological, utilitarian, virtue-ethics, feminist, fundamentalist, Nietzschean and emotivist approaches to morality.

2.6. KOHLBERGIAN TO NEO-KOHLBERGIAN RESEARCH: DEVELOPMENT AND CRITIQUES

The Minnesota group saw its work as an advancement of Kohlberg’s work, not a rejection of it, clearly identifying their own research as “Neo-Kohlbergian” (1999b, 2000). Eager to point out the many aspects of development in which their work agreed with Kohlbergian theory, the group identified several important facets of Kohlberg’s

work that persisted in this Neo-Kohlbergian framework. Like Kohlbergian research, the group's work with the DIT focused on cognition and the personal construction of categories that scaffold the cognitional aspects of moral reasoning and decision-making. Indeed, the group has been criticized, as Kohlberg was, for focusing too narrowly on cognition largely to the exclusion of emotional aspects of moral reasoning. The group also agreed with Kohlberg's basic assertion that moral growth is part of an individual's attempt to make sense of experiences, particularly those of a social nature. Finally, like Kohlberg, the group conceived of moral growth as a forward movement toward higher integrations of moral understandings, though, as mentioned above, the group came to envisage moral growth in "soft stages" rather than the step or staircase model Kohlberg favored. However, the group's work benefitted from reflection upon the many practical and theoretical critiques of Kohlbergian theory and utilized advancements in the fields of Philosophy and Psychology to hone the horizon of their research.

Critiques of Kohlbergian theory have come in many forms and offer many insights for neo-Kohlbergian research. Kohlbergian theory's use of and dependence on particular foundations elicited two major lines of criticism: first, a basic rejection of the normativity-based philosophical foundations of Rawlsian, Kantian and deontological ethics; and, second, a claim of bias, particularly gender and cultural bias, embedded in the foundations of the work done with these theories. However, Rest and his colleagues sensed that they could adequately address the majority of those criticisms while maintaining the best of Kohlberg's insights.

2.6.A. Critiques of Foundationalism

Kohlberg's theories came under fire from many sides but it has been late 20th century developments in moral philosophy and psychology that have brought into question the validity of its philosophical foundations (Maxwell, 2010; Rest et al., 2000). Contemporary philosophy's rejection of a singular principle or "grand narrative" to which we might appeal as a foundation for claims about moral judgment or morality provides an important criticism of theories like Kohlberg's which employed a unifying principle as a means to securing a comprehensive standard for measurement. Kohlberg made significant use of Rawlsian "justice operations," including notions of reciprocity and ideal states of fairness or equality, with great value placed on reversibility (as discussed in Rawls' later work [2001] under the auspices of the so-called "veil of ignorance") in developing his MJI and its scoring modes. These Rawlsian notions in turn relied heavily on Kantian and deontological conceptions of moral duty and rectitude. Though Kohlberg claimed to have purged his theory of philosophical content, preferring a more abstract framework that might distil pure cognitional structures, critics claimed that he deductively utilized a foundational philosophical principle which was controversial and overly directive in his assessment tools (Thoma, 2002).

Rest largely rejected these philosophical underpinnings of Kohlberg's work (1999b, 2000), adverting to the weakness of depending on top-down, abstract principles and instead chose to work from specific cases toward an agreed upon or "common morality" that emerges from a community's reflective consensus on moral issues (Rest, Narvaez, Thoma, & Bebeau, 2000). Making use of the insights of a social constructivist approach, utilizing "looser, broader notion[s] of cognitive advance" (Rest et al., 2000, p.388) and building the DIT via a "bottom-up" mode, Rest and his colleagues established

a tool for activating moral schemas that resonate with subjects. Thus, the Neo-Kohlbergian stance involves the claims that justice is not found *in* individuals but is the fruit of lived, communal meanings that are interpreted and recognized by individuals who participate in those communal understandings and go on to reflect on the ideals and logical coherence of those values. From this point of view, notions of what is moral and just are the result of incremental, non-arbitrary social cooperation that values impartiality, organization of rights and responsibilities, and balancing of self and social concerns.

2.6.B. Gender Matters

Though critiques of Kohlberg's work and Kohlbergian theory come from many different perspectives, the most well-known, even in popular spheres, was the claim of gender bias leveled by Kohlberg's student and colleague, Carol Gilligan (1982). Gilligan contended that Kohlberg's studies and conclusions excluded women's ways of approaching and reconciling moral problems from the higher stages. The issue of gender bias presents a two-fold problem, including the charge that the principle on which the stage theory hinges is at root biased against females, and that Kohlberg's method of data collection was severely flawed. Gilligan's claims were challenged by later studies (Bebeau & Brabeck, 1987; Lifton, 1985; Nunner-Winkler, 1984; Walker, 1984) but their impact was significant.

In 1982, Gilligan published what became a very popularly-known critique of androcentric bias in contemporary psychological theory, particularly as regards moral development theory, including Kohlberg's work. Noting Nancy Chodrow's (1978) work in examining the separation and individuation work of gender identity development, Gilligan points out that empathy and attachment produce an ethos of care and relationship

in girls, whose identity work is primarily accomplished via an experience of being *like* their mothers, ie. sharing an identity with the mother who is and who symbolizes the primary giver of care, as opposed to boys whose main identity work is done in the context of separation and differentiation from their mothers. Gilligan goes on to postulate that female identity development is so tied to modes of attachment that the separation and detachment of male development seem threatening and problematic and are thus not valued or highly integrated in women's ways of approaching conflicts. Gilligan's analysis of interviewees' responses to the famous Heinz dilemma of Kohlberg's MJI test highlights the different approaches of male and female subjects along these lines. In Gilligan's study, the female subject seeks relational approaches to the problem, asking for more options than the MJI interviewer offers, and attempting to find a resolution that is inclusive of the complex needs of all the parties in the dilemma, while the male subject, construes the moral problem as a question of rights and attempts to construct a logic of justice by ordering the claims of the characters in the dilemma. Against the backdrop of a justice-operations based theory, like Kohlberg's, this male approach gives a score that reflects a higher developmental stage. Gilligan's own study of 29 women considering abortion construes a female mode of moral reasoning which moves through three increasingly complex iterations of the relationship between self and others, the transitions of which pivot upon "critical reinterpretation of the conflict between selfishness and responsibility" (Gilligan, 1982, 105), articulated in a distinct moral language that Gilligan claims is eluded in Kohlbergian stage analysis.

Recent reviews of the literature examining gender differences in moral reasoning show remarkably little or mixed findings along the lines of Gilligan's claims (Bebeau &

Brabeck, 1987; Bruess & Pearson, 2002; Mayhew, 2010; Walker, 2010). Meta-analyses reported by Walker (1984, 1986, 2010), Thoma (1986), and Baumrind (1986) and Bebeau and Brabeck (1987) show that gender difference yields little variance in MJJ results of more than 10,000 subjects and reveals that at every age and education level, women score higher than men in DIT testing of moral reasoning. Mayhew's meta-analysis of 43 studies (2010) that examines the relationship between gender and moral reasoning via the DIT, observes that just over half of the studies find women utilizing more sophisticated strategies in approaching moral dilemmas, while the remaining studies show no difference or found men to score higher along developmental lines. Moreover, Thoma's meta-analysis highlights the finding that in samples of over 6000 subjects, education was more than 500 times more powerfully predictive of moral reasoning than gender. To gain a better understanding of the disparity between Gilligan's claims and the findings of many researchers, Bebeau and Brabeck examined a variety of studies that take up Gilligan's claims from several different perspectives (1987). Though studies show little difference in MJJ and DIT measures of moral reasoning along gender lines, Bebeau and Brabeck's meta-analysis goes on to consider a variety of claims of flaws in Kohlbergian and neo-Kohlbergian stimulus material (i.e. terms of dilemmas as well as details about characters and situations) and scoring rubrics that emphasize or even exclusively posit justice operations over and against care orientations. In research done using Kohlbergian dilemmas and seeking both justice and care orientations, they find no overall stage difference between genders and observe that men and women are equally likely to demonstrate care and justice orientations in their moral reasoning. Their analysis of research suggests that moral orientation is not determined by gender but by the type of

dilemma one considers oneself to be confronting, which in turn lead Bebeau and Brabeck to hypothesize that it is in the realm of moral sensitivity, the very construal of what makes a dilemma a dilemma, where males and females diverge.

In a meta-analysis of studies of gender difference in Kohlbergian measures of moral reasoning, Walker (1984) also takes up the claim of gender bias in Kohlberg's work in moral development theory, exploring three basic issues of this claim: first, that Kohlberg's maleness itself brings a bias to the work, Walker points out that a number of Kohlberg's colleagues who shared significantly in his work were women and a senior author of the revised scoring methods of his MJI was female; second that Kohlberg's sample included only men. Walker points out that relatively little data support a claim that women do not follow Kohlberg's stages and that studies in subsequent years largely demonstrate no significant difference between men's and women's attainment of higher stages of moral reasoning (Armon & Dawson, 1997; Bebeau & Brabeck, 1987; Walker, 1984, 1995). Third, the predominance of male protagonists in the dilemmas used in the Kohlberg's study biases the results. For these reasons, Walker points out that studies are equivocal in findings with same-gender and opposite gender protagonists and subjects (1984).

2.6.C. Critiques of the DIT as a Measurement Tool: What Does It Measure?

Beyond these more well-known criticisms of Kohlbergian and Neo-Kohlbergian theory, a number of other criticisms emerged as the DIT became a popular and trusted research tool in the 90s and the first decade of the 21st century. During the course of the DIT's development, many researchers including Kohlberg himself questioned the Minnesota group's assertion of the measurement tool's strength and reliability (Elm &

Weber, 1994; Emler, Palmer-Canton & St. James, 1998; Kay, 1982; Murk & Addleman, 1992; Rest, 1979). Kay asserted that methodological and conceptual difficulties associated with the test limited its ability to overcome basic conceptual inadequacies of Kohlbergian moral stage development theory generally, noting that the verbal and literary designs of both the MJJ and the DIT exclude pre-adolescents and confound reasoning or comprehension skills or other intellectual development with moral development. Other critics have suggested similar confounding variables or found evidence to suggest that moral reasoning results reduce to political or religious orientation (Emler, Resnick & Malone, 1983, Emler et al., 1998; Getz, 1984). Murk and Addleman (1992) raise important questions about religiosity as an important variable in the advancement of moral reasoning and examine significant correlations between DIT scores and a set of five variables, including age, educational attainment, religious affiliation, gender, and Internal-External Locus of Control Scale scores. Though Murk and Addleman's findings were less impactful than Thoma's and Rest's, their findings supported Thoma's and Rest's claims that age and education, though often confounded, account for much of the advancement in moral reasoning. Rest and Thoma's research suggests that age accounts for as much as 38 to 52 percent of variance and Thoma's finding demonstrates even greater percentages of variance attributed to educational attainment (Rest & Thoma, 1985, Rest 1986; Thoma, 1986). Their findings are supported Colby and Kohlberg's 1987 research claims of a strong relationship between cognitive variables and moral judgment and King and Mayhew's meta-analysis finding that among 45 studies using design strategies that directly test the effects on moral reasoning of participation in formal higher

education, 90% showed a significant relationship between formal education and the development of moral reasoning.

Finally, Elm and Weber (1994) concur with Kay's criticism of the extensive use of quasi-experimental studies in assessing the DIT as a fit measure. While these issues will be addressed in a later chapter of this work, it is important to note Kay's subsequent hypothesis that the DIT confounds variables like educational achievement, social values and intellectual ability with moral development, and is thus a measurement of these variables rather than of a unique developmental trajectory (Kay, 1982). These critics join a number of researchers and theorists over the years who have speculated about the adequacy of the MJT and DIT in capturing unique aspects of moral development and their capacity to evaluate any correlation between particular interventions and advancements in moral reasoning. Researchers have long wondered about the efficacy of moral development measures and the impactful and confounding effects on moral development of variables (besides gender, mentioned above), such as educational attainment or intellectual achievement (Burwell, Butman, & Van Wicklin, 1992; Mentkowski, 1983, Mentkowski et al., 2000), educational environments (King & Mayhew, 2002; Mayhew, Fernandez, & Deluca, 2007; Mayhew, Seifert & Pascarella, 2010; Mayhew, Wolniak & Pascarella, 2008; McNeel, 1991, 1994; Maeda, Thoma, & Bebeau, 2009; Pascarella & Terenzini 2005) verbal ability (Thoma, Narvaez, Rest & Derryberry, 1999), stages of identity development (Bruess & Pearson, 2000), political identity (Emler et al., 1983; Frimer, Biesanz, Walker & MacKinlay, 2013; Thoma, Barnett, Rest & Narvaez, 1999; Thoma, Narvaez, Rest & Derryberry, 1999), socioeconomic status (Finger, Borduin & Baumstark, 1992; Mentkowski & Strait, 1983; Rest 1979) religious, cultural or socio-

political ideology (Murk & Addleman, 1992; Narvaez, Getz, Rest, & Thoma, 1999), age (Armstrong, 1993; Shaub, 1994), race and ethnicity (King & Mayhew, 2002; Murk & Addleman, 1992). Of these variables, it is noted that the persistent consideration of the effects of verbal ability and political orientation on DIT scores has largely emerged in the context of questions about the construct validity of the DIT and the Minnesota group's responses (Emler, Palmer-Canton, & St. James, 1997; Mayhew, Seifert & Pascarella, 2010; Narvaez, Getz, Rest & Thoma, 1999). Finally, synthesizing over 200 empirically based studies, researchers noted that cognitive motivation, a willingness on the part of an individual to engage in effortful thinking, was associated with moral reasoning development (Cacioppo, Perry & Kao, 1984; King & Mayhew, 2002; Mayhew & King, 2008; Mayhew, Seifert & Pascarella, 2010; Mayhew, Wolniak & Pascarella, 2008). King and Mayhew's 2002 meta-analysis concurs with Rest's 1999 finding that formal education attainment is "by far the most powerful demographic correlate of DIT P-scores, typically accounting for 30 to 50 percent of the variance in large, heterogeneous samples (p.70). In 2010 review of literature, Mayhew, Seifert and Pascarella note that within quasi-experimental and correlation-based longitudinal designed research, formal college attendance is consistently demonstrated to promote increased moral reasoning, quite apart from the gains attributed to general maturation or age.

Regarding the impact of variables on DIT scores and moral reasoning development theory, the Minnesota group has argued forcefully and regularly that its view was not that the DIT provided a *pure* measure of moral reasoning or moral development, nor that it seeks to exclude other variables in evaluating this type of development. The group contends for instance that moral reasoning cannot be reduced to

cultural ideology but that the process of cultural socialization interacts with the cognitive construction of meanings in such a way to create *moral thinking*—the ability of a subject to assess the moral or ethical aspects of situations, the capacity to reflectively discern right and wrong behaviors and attitudes and to provide coherent rationales for that thinking (Narvaez, Getz, Rest, & Thoma, 1999; Thoma, Barnett, Rest & Narvaez, 1999). In this way, the Minnesota group seeks to make sense via the DIT of the work individuals do in pivoting from the realm of moral sensitivity (Component 1), the recognition of a moral aspect of a situation, through moral reasoning (Component 2), the determining and defining of right or ideal action, toward the selection and execution of moral action (Components 3 and 4). Rest contends that DIT research adequately examines development of this second component and asserts that while moral reasoning development does not reduce to specific variables, much work needs to be done to determine the impact of various aspects of individuals' experiences and educational opportunities on this development (Mayhew & King, 2008; Rest et al., 1999).

2.7. RECENT RESEARCH WITH THE DIT: EDUCATIONAL AND PEDAGOGICAL

INTERVENTIONS IN MORAL DEVELOPMENT RESEARCH

In an attempt to examine both college student moral development patterns and the usefulness of DIT research in determining and identifying practices that advance or foster moral growth, extensive research has emerged in the past decade examining pedagogical practices and educational interventions which might precipitate moral development. Using insights from Neo-Kohlbergian theory and practice, educational researcher Mathew Mayhew has investigated many important areas of curricular conditions and educational intervention models that might impact and advance moral reasoning.

Mayhew and Patricia King argue that in the context of the demands of democratic society, themes of ethical and social responsibility are not simply a retrieval of the missions of early American colleges and universities but are applicable today to the missions of religiously affiliated and secular institutions alike (King, 2009; Mayhew & King, 2002, 2008; Mayhew, Wolniak & Pascarella, 2008; Reuben, 1996). Mayhew and King posit moral reasoning as a key characteristic that grounds the goals of democratic society and advert to the helpfulness of Neo-Kohlbergian developments in thinking about how educational institutions might begin to address these mandates. They also note three decades' worth of DIT data which overwhelmingly relates the development of moral reasoning to participation in formal higher education. The confluence of renewed interest in promoting moral development and increased confidence in the DIT's reliability has convinced Mayhew and his colleagues to pursue a closer examination of the specific aspects of higher education that affect moral reasoning development.

Mayhew and King note that purposeful educational interventions aimed at fostering moral development and moral reasoning in students fall broadly into two categories: the first involves use of specific *content*, while the second emphasizes pedagogical *strategies* (2008). In the case of content designed to stimulate moral reasoning, Mayhew and King identify two modes used in most classroom interventions: *explicit* moral content and *implicit* moral content modes. In an explicit mode, instructors are apt to teach principles of ethics, model advanced reasoning and perspective-taking skills, and engage students in considering morally challenging dilemmas in the hopes of encouraging increased capacity for alternative perspective-taking and recognition of moral complexity. In the second, implicit mode of content-driven interventions,

instructors tend to utilize broader theories of social justice and use social issues as prompts for discussions about social ills such as oppression, power, privilege, racism, sexism, etc. In brief, Mayhew and King present the major difference between these modes as one of focus. In the explicit mode, the logic of moral principles and the nature of the moral dilemmas are highlighted, while in the implicit mode, though a need for higher level moral reasoning is also demonstrated, harmonious social relations are emphasized. A second category of educational intervention overviewed by Mayhew and King is that of pedagogical strategy, in which the impact of a variety of educational activities such as role-taking, service-learning opportunities, perspective-taking discussions, and cognitive-disequilibrium assignments are examined.

Within both categories of educational intervention, Mayhew and King find inconclusive research evidence of particular course effects on moral reasoning, in accord with Rest's 1979 mixed results on short-term educational interventions. In a review of over 500 studies conducted in the past thirty years, Mayhew and King (2002) find surprisingly little evidence attributing growth in moral reasoning to particular courses, pedagogical styles, or educational interventions. Further, they note that among over 60 of these studies observing the effectiveness of course-related interventions, most target graduate students in professional programs such as medical/dental, accounting or law programs, or undergraduates enrolled in upper division, pre-professional programs, despite findings of significant effects of the first year in college on moral reasoning growth (2002). These researchers thus identify a gap between the clear evidence of undergraduate, particularly first year, advancement in moral reasoning and the dearth of explanatory evidence regarding *exactly what it is* in formal undergraduate higher

education that precipitates moral growth. Mayhew has gone on to explore a number of aspects of higher education that may help us move toward an answer, including particularly illustrative work analyzing data from a large, multi-institutional study (Mayhew & King 2008; Mayhew, Seifert, & Pascarella, 2012) and a longitudinal study of students enrolled in courses with differing modes of moral content. Mayhew and King conclude that including explicit moral content in a course appears to be impactful in fostering development of moral reasoning and speculate that explicit moral content may offer students a helpful language and set of structures to utilize when approaching difficult moral challenges. Studies like this are a first step in identifying more precisely what it is in higher education that precipitates moral growth and moral reasoning.

In more recent and very important studies, Mayhew and his colleagues observed DIT data from a sample of 1,469 first year, full time students from 19 two- and four- year colleges, mostly within the liberal arts tradition, who participated in the large-scale, longitudinal 2006 Wabash National Study of Liberal Arts Education (WNS) (Mayhew, Seifert, & Pascarella, 2010, 2012). The sample selection reflected the researchers' goal of a diversity of institutional size, location, type residence pattern. Via a series of factor analyses, Mayhew examined a set of demographic, course-taking behaviors, educational practices, and co-curricular variables, as correlated to the development of moral reasoning. Interestingly, though the effects of co-curricular and classroom experiences taken together netted a significant but small effect in advancing the moral reasoning capacity of these first year students, of four variables used regarding course-taking behavior, the variable that yielded a significant effect on moral reasoning was the extent to which their courses helped them understand historical, political and social connections

of past events and brought these into dialogue with students' own situations. These findings lead researchers to the conclusion that

exposing students to curricular content that engages them in critical dialogue with the past [that] may also encourage them to situate themselves, their ideologies, and their notions of fairness, in the larger, meta-narrative of human history; such an expanded paradigm for understanding how self is related to other is a hallmark of advanced moral reasoning (Mayhew, Seifert, & Pascarella, 2010, p. 379).

In addition, findings of the study recommend frequent positive challenges to students in the classroom setting, encouraging the application of course content to actual problems, and teaching critical skills needed to point out and correct false arguments in basic and communal points of view. The study also finds that quality of teaching and interactions with faculty outside the classroom are significantly linked to advances in moral reasoning, especially for first year students, suggesting that in terms of moral reasoning development, interpersonal connection in the classroom is as important as what is taught. Findings in the study identify moral reasoning as a distinctive area of inquiry, related to but not reducible to political orientation, gender or racial biases, or other cognitive constructs such as intellectual, verbal or academic abilities or cognitive motivation. Mayhew and his colleagues claim that the implications of this study are far reaching for higher education, noting that very few studies have as yet attempted to “unpack the collegiate experience” to discover precisely which aspects of this experience significantly affect the sort of moral development called for by college and university, state and federal governing bodies.

Halliday and Frantis (2006) found similar data in their study on the usefulness of ethics courses for undergraduates enrolled in a health care program, noting the importance of classroom practices that include many practical examples in demonstrating theoretical foundations for ethics programs. Halliday and Frantis pointed out that structuring classroom discussions around moral theory and moral issues means that at some points in a course, teachers will be asking students to think at a developmental level beyond their own capacity. This “challenge and support” model, however, is precisely what Mayhew seeks in classroom practice, offering challenges to lower order moral reasoning, providing space for discussion that invites puzzling through inadequate or insufficient moral paradigms, and presenting more inclusive, more complex and more nuanced ways of bringing ethical and moral theory into dialogue with concrete and meaningful contemporary issues. These practices are certainly suited to the developmental trajectories of Piagetan, Kohlbergian, and Neo-Kohlbergian models of moral reasoning.

Indeed, it is consideration of this practice that lead Mayhew, Seifert and Pascarella (2012) to reanalyze their WNS findings in terms of yet another factor of students’ moral development. In this analysis, Mayhew and his colleagues examined information from the DIT which situates students either in a *consolidation* or *transition* phase of moral reasoning, adverting to the impact of the stability of a subject’s moral positioning on her openness to educational interventions. Using the DIT’s N2 scoring, these researchers were able to distinguish students who are apt to use consistent and independent cognitive strategies in facing a moral dilemma (consolidated phase) from those utilizing a variety of cognitive strategies and prefer to use situational and contextual

cues in decision-making (transition phase). Using Astin's Input-Environment-Output model, Mayhew, Seifert, and Pascarella sought "to deconstruct the college experience into those curricular, co-curricular, and teaching practices potentially responsible" (2012, p. 24) for the moral development gains typically found in studies of college students.

2.8. IMPLICATIONS FOR LIBERAL EDUCATION AND THE BOSTON COLLEGE

PERSPECTIVES PROGRAM

Studies on the impact of educational interventions on moral reasoning regularly examine the type of course content and practices that are mainstays of liberal education. But participation in liberal education and humanities programs is in steep decline in American higher education, particularly within research universities and large multiversities, where waning numbers of humanities majors evinces higher education consumers' profound ambivalence toward these traditional educational formats. Against the backdrop of a challenging economic reality, parents and students are opting for pre-professional and vocational programs in increasing numbers. The 2012 CIRP Freshman Survey reported an all-time high of nearly 88% of incoming freshmen identifying "to get a better job" as their top reason for attending college, a reason which has topped the list of reasons for college participation since 2006. Not surprisingly, the numbers of students majoring in humanities has been steadily declining since the 1970's, with business, health professions, biological sciences, and engineering occupying the top four intended fields of study of CIRP respondents. As a recent New York Times article points out, a university with a long tradition of excellence in the humanities like Stanford now finds itself with only 15% of its students majoring in humanities which account for 45% of the faculty (Lewin, 2013). As fewer and fewer students choose liberal arts and humanities

majors, the relevance of liberal education, as well as its central aims, are no longer a given.

Liberal education faces formidable challenges within the academy as well. Liberal arts education as we know it today emerged in the early decades of the twentieth century as a humanist retrieval of the classical liberal arts curricula. Its champions envision it as a response to the increased specialization and value-neutral philosophy of the research university model and commend its ability to revitalize the moral purpose of higher education. Liberal education claims to offer a timeless curriculum – one that articulates perennial questions and expresses transcultural ideas – and to have revived the pursuit of moral character and values as legitimate concerns of a college education (Reuben, 1996). But the aims of liberal arts education are notoriously difficult to assess and thus don't square easily in the present research-dominated higher education scene. Liberal education focuses on critical thinking, ethics, interdisciplinarity and critical/cultural analyses, utilizing pedagogical strategies and classroom practices such as small class discussions, course-related service-learning, multiple-perspective taking, cross-disciplinary study and articulation of basic principles of the common good and social justice. Seeking to develop the “whole person” through a diverse, humanities-based curriculum, liberal education pursues the development of character and ethics, the advancement of students' critical thinking across a broad range of fields, and an engagement with deeply rooted intellectual inquiries beyond the parameters of practical and professional concerns (Cox, 1985; Hirst, 1965; Pascarella, Wolniak, Seifert, Cruce & Blaich, 2010). It privileges the transmission of knowledge over the production of knowledge, seeking primarily to cultivate in students what the 1828 Yale Report refers to as the disciplines and power of mind, habits

and skills of dexterous thinking, and a balance of character (Brubacher & Rudy, 1997). These aims, however, are difficult to operationalize, evaluate and assess, eluding many social science research tools. Nineteenth and 20th century paradigmatic shifts in epistemology and the subsequent dominance of empirically based research have left liberal education with limited means of demonstrating its achievements or outcomes. Further, liberal education faces important and daunting postmodern critiques of normativity claims that attempt to standardize or objectify moral development. At present, liberal education needs to justify its place in the academy as it never has before and liberal education programs and majors must prove their worth in the public and academic scene (Cox, 1985; Fish, 2003a, 2003b; Reuben, 1996).

In the spirit of that inquiry, this study sought to explore if and to what extent the moral reasoning capacity of first year college students is positively affected by the cornerstone course of the Perspectives Program, an implicit goal of which is the promotion of moral development in students. The study employed a secondary analysis of a university-sponsored assessment of the course, which included over three hundred pre- and post-test student surveys and an analysis of student essays on the impact of the course. DIT data and essays were examined in a mixed-methods analysis of the course's influence on students' moral reasoning development in the hopes of contributing to ongoing research in student moral development and the advancement of educational strategies that promote such development.

Chapter 3:

Research Intervention and Quantitative Methodology

3.1 INTRODUCTION: RATIONALE OF THE RESEARCH QUESTIONS AND DESIGN

Though research indicates that college participation positively influences student development, much work remains to be done to identify which facets of student development are most impacted by college participation and which aspects of college are influential in advancing these developments. In their 2005 meta-analysis of over 2,600 studies, Pascarella and Terenzini confirmed the impact of college participation on college students, noting net and long-term developmental effects. In the case of moral development, this meta-analysis found significant positive changes in moral reasoning in college students from freshman to senior year in research that allowed controlling for a number of factors including subject maturation, socioeconomic status, cultural context, and levels of precollege moral reasoning and intelligence. Moreover, this finding holds across measurement instruments. Despite evidence that college provides a uniquely fruitful context for growth in moral reasoning, researchers admit that controlling for all variables which might impact this growth is a major challenge. However, research strongly suggests that 1) moral growth in college is among the most significant developmental advances of college students, 2) that these advances persist after college, and 3) that this growth does not seem to reduce to other factors.

Pascarella and Terenzini (2005) note that recent quasi-experimental evidence suggests that general education and liberal arts education programs that integrate

instruction in philosophical methods of analysis in ethics and ethical decision-making may enhance moral growth, while the effect of ethics courses and ethics interventions generally give mixed results and little evidence is found to support that major field of study has any effect on moral development. Pedagogical interventions offer similarly mixed results, with service learning requirements not seeming to impact principled reasoning unless combined with course reflection within course content, while coursework involving role-playing dilemmas or moral dilemma discussions are found to be impactful. It is clear to many who are interested in exploring the impact of educational interventions that much work needs to be done to specify what sorts of programs and pedagogical strategies enhance and advance development.

The present study was stimulated by these questions and concerns and addresses some of them through an examination of assessment data of a course with an implicit goal of enhancing students' moral development. Two chapters outline the mixed-methods approach to the study. Chapter Three begins with a sketch of the research questions and design of the study and gives a rationale for its quasi-experimental nature. A thorough explanation of the research site and the course which serves as the intervention at the center of the study is also offered in Chapter Three, as well as a summary of the data collection procedures. The chapter commences with a detailed explanation of the quantitative measurement tool used in the assessment, the Defining Issues Test (DIT). Chapter Four reviews the development of the qualitative component of the study, offering a detailed account of the development of writing rubrics used to analyze student essays and an explanation of how the quantitative and qualitative data were synthesized. Chapter Four will also include an analysis of the limitations of the study and its methodologies.

3.2 THE RESEARCH QUESTIONS

This study sought to contribute to discussions regarding the impact of college participation and the influence of curricular interventions on moral reasoning as a proxy for moral development through a secondary analysis of assessment data from a liberal education course with an implicit goal of promoting moral development. The study posed three research questions:

1. Does the moral reasoning capacity of first year college students increase in a Great Books course with an implicit goal of promoting moral development?
2. How do students perceive the moral dimensions of the course?
3. To what extent is the development of students' moral sensitivity and moral reasoning evidenced in their own written reflections about the importance of the course?

This investigation involved a mixed-methods approach and thus presented two types of research data: 1) an analysis of quantitative data from pre- and post-test surveys measuring change within students moral reasoning and comparison of their reasoning levels and growth with national student norms, and 2) an analysis of qualitative data from open-ended pre- and post-intervention essays assigned at the beginning and end of the course via two analytic rubrics developed by the primary researcher. The rubrics examined students' self-reported perceptions of the moral dimensions of course content and student language patterning that evinces moral sensitivity and moral reasoning development.

3.3 THE RESEARCH DESIGN

The study consisted in a secondary analysis of course assessment and was thus quasi-experimental in nature, since there was no random assignment to the educational intervention, which in this case was participation in the course. Students chose to participate in the course within their first year of college. The course offered students one way to complete a set of required courses within the liberal arts “core” program but the course itself was not required by the university. Self-selection of students into this course may be one of several important confounding factors that impact the internal validity of the study’s findings. A good deal of moral development research literature addresses these types of confounding factors, all of which will be considered in Chapter Five’s articulation of the findings of this study. Because the course was one of only a handful of courses intentionally designed to be a year-long course, the study suffers from not including a suitable control group. However, the data examined in this study were designed as a larger course assessment and were thus not intended to be measured against a control group. How this design element impacts the findings of this study is discussed in Chapter Four’s section on the limitations of the study design.

As regards the first of the central questions of the study, the research design assumed a null hypothesis, ie. that students would experience no significant gains in moral reasoning development during the year. Literature shows clearly that normal maturation and college participation generally impact student moral development, though some research suggests that the greatest gains in moral reasoning are found in the second year of college. Thus, some gains were expected. Findings were generally measured against national benchmarking and trends. With respect to the second and third research questions, the study sought to examine students’ own perceptions of the moral

dimensions of the course and course content, and to observe students' own expressions of their developmental gains.

The research design employed a mixed-methods, concurrent triangulation (sometimes referred to as “simultaneous triangulation”) approach to assessment materials which came in both quantitative and qualitative forms. Quantitative and qualitative data collection was implemented concurrently during the period of the course assessment, ie. the DIT and essays were administered at the same points in the academic year (September and April, 2012), though these tasks were not explicitly linked by instructors in the classes. Triangulation is an apt method for examining phenomena like moral development and its relation to an educational intervention since this methodology “may be used not only to examine the same phenomenon from multiple perspectives but also to enrich our understanding by allowing for new and deeper dimensions to emerge” (Jick, 1979). Morse points out in her overview of methodological triangulation that deductive projects working with a priori frameworks are best designed with quantitative data taking precedence, complemented by qualitative data (1991). Thus analysis was completed sequentially, beginning with an examination of DIT data and subsequent categorizing of subjects based on scored outcomes (eg. low scorers with significant gains, low scorers with limited gains, high scorers with significant gains, etc.), followed by an analysis of student essays grouped as such.

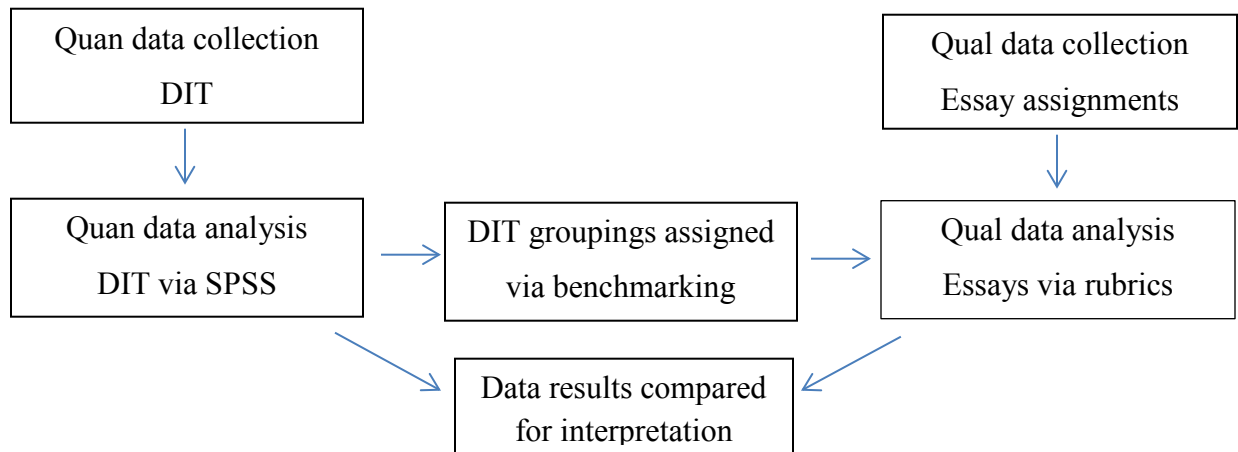
Mining student writing in the analysis stage was intended not to corroborate quantitative findings, but rather to identify aspects of the course that were associated with various types of moral growth shown within DIT data sets. Student self-reported impressions of moral dimensions of the course helped to elaborate on various DIT

findings in ways that deepened and enriched those findings rather than simply validating DIT data (Greene, Caracelli, & Graham, 1989). Integration of quantitative and qualitative data was completed during both the data analysis and interpretation stages of the research process. Qualitative material was then analyzed to identify motifs and themes from student writing and to capture a holistic and coherent depiction of subjects' moral development (Creswell, Plano Clark, Gutmann & Hanson, 2008; Jick, 1979).

There were several options for merging the study's mixed data (findings from the DIT and essay analysis). Qualitative data could have been considered first in an exploratory mode, followed by quantitative data that might or might not validate the findings of the qualitative data. A second option was to analyze the large sample of quantitative data first and then use smaller cases within the qualitative data to expound and illustrate the kinds of insights offered by the larger data set. This second model refrains from prioritizing quantitative data, offering qualitative data as a way to identify patterns within grouped DIT levels and profiles as well as providing a larger sense-making of the impact of the course. The second model was better suited to the original purposes of the course assessment, since that process sought to uncover what sorts of development might be connected with participation in the course and which aspects of the course were impactful to student development. Preceding student writing analysis (qualitative data) with DIT data analysis (quantitative data) allowed the larger swath of quantitative information to provide a general picture of subjects' moral development and in particular of subjects' gains in moral reasoning, while the second phase of essay analysis offered an opportunity to drill down into what students themselves reported as

pertinent to advancement and growth in these areas. The concurrent triangulation design of the study is illustrated in Table 3.1.

Table 3.1 Concurrent Triangulation Design



3.4 THE RESEARCH SITE AND TREATMENT/INTERVENTION

3.4.A. The Research Site: Boston College

The educational intervention in this study was a year-long course in the great books tradition offered to first year students at Boston College, a religiously affiliated, mid-sized research university. The school is part of a 28-college network of Jesuit, Catholic colleges and is a strongly mission-driven institution. As such, it has a long history of educating students in the liberal education tradition. That tradition is attached not only to the original aims of the university which was founded in 1863, but also to the educational traditions of Jesuit education, as formulated in the 16th century *Ratio Studiorum* (“plan of studies”) which serves as the de facto official blueprint for Jesuit higher education. Also connected to this tradition is a strong commitment to “whole person education,” understood by the college as a responsibility to attend to the

integration of the spiritual and social aspects of students' lives in addition to their intellectual development. As such, the education of a student's character and attention to the moral development of students is part and parcel of the institution's aims of "student formation." A document recently produced by the university, "The Journey into Adulthood" explains this notion of educational formation this way: "[i]nseparable from...intellectual formation is the goal of shaping of character, of producing graduates who will take seriously the challenge of living good lives and making the world a better place" (Appleyard, 2008).

The college's national ranking within the top 40 universities in the US (as reported in several major, national rankings), a wide array of liberal arts and pre-professional fields of undergraduate and graduate study and its location close to a popular US urban area make it a popular choice for applicants. The school regularly receives over 20,000 applicants each year for placement in a freshman class of 2,250, at an acceptance rate of 32%, with 82% of freshmen having been in the top 10% of their high school classes. Boston College regularly highlights its commitment to liberal arts curricula and to the moral aims of liberal education. Thus, one can assume that participants in this study were fairly high-achieving, motivated students who had competed to gain a spot in a highly selective institution which values intellectual excellence as well as personal, emotional, social and spiritual flourishing.

Data from the National Survey of Student Engagement (NSSE) Jesuit Institution consortium survey (a NSSE supplement uniquely designed to capture aspects of the Jesuit educational experience) offer helpful considerations in understanding the subjects of this present study (Boston College Office of Institutional Research, Planning and Assessment,

2013). Within its Enriching Educational Experiences section the NSSE survey, seventy-nine percent of Boston College seniors who completed the 2013 NSSE reported that the university contributed “quite a bit” or “very much” their own development of or clarification of a personal code of values/ethics. This finding was statistically significantly higher than their counterparts at other Jesuit colleges, other colleges and universities within the same Carnegie classification, and all institutions participating in the NSSE survey. Both freshmen and seniors that same year reported higher indicators of reflective and integrative learning than their counterparts at other colleges as well, statistically higher than other Carnegie class and NSSE institutions.

Thus, the research site and its student demographics presented both opportunities and challenges to this analysis of student moral development. On one hand, Boston College students do not comprise a typical young adult or even college student sample. They represent a selective and somewhat elite group of students from predominantly privileged backgrounds attending a very competitive school the aims of which explicitly include moral and ethical development. Moreover, social justice initiatives and volunteerism permeate the campus culture, evidenced both by NSSE data and by self-reported, robust competition among students for positions in volunteer programs and service and immersion trips. It was thus helpful in this study to examine moral development data not merely with an eye to national and age-related benchmarks, but to focus on specific types of DIT shifts (such as low scores to high, high scores to higher, high scores to low, etc.) of individuals. In other words, some students would be expected to begin with high moral reasoning scores relative to other young adults in their age cohort but would not make significant strides within the first year relative to those

initially high scores. Hence, the quantitative component of this study sought to identify patterns of students' moral development, comparing scores with national trends, and to establish groups of students within the cohort whose scores reflect significant moral development gains (demonstrating gains of one standard deviation or more) from low to high scores or from high to higher scores, as well as those who exhibited significant losses in moral development (demonstrating losses of one standard deviation or more). The qualitative component of the study examined student writing to uncover course-related insights into students' moral development gains and losses. An examination of student writing within and across these groups sought to reveal how students use, perceive, and were impacted by coursework in their own development.

3.4.B. The Treatment/Intervention and Sample: "Perspectives in Western Culture"

The intervention that served as the independent variable of the study was a course entitled *Perspectives in Western Culture*. It is the first, cornerstone course within a comprehensive, interdisciplinary program of study housed in the university's philosophy department. The course is widely regarded by students, faculty and administrators as the type of course that exemplifies the liberal education tradition of the university. Nineteen sections of the year-long (2 semester) course are offered each year to freshmen at Boston College, with a limit of 25 students per class. In 2013, 452 out of a total freshman class of 2,405 (n= 1,286 women and 1,119 men) registered for the course, approximately 20% of the college's freshmen. It is important to note that getting into the program is sometimes challenging. Many students who would like to participate in the program are not able to enroll due to the popularity of the course. Incoming freshmen are informed about the course during their freshman orientation sessions and via a course catalog sent out in the

summer months before they arrive. Students register for the course during freshman orientation programs, with 3-4 seats in each class becoming available during the 7 orientation programs held during the summer months. Student orientation leader and academic advisor training includes a detailed description of the course's interdisciplinary nature as well as its rigorous workload. Students are often alerted to the fact that as a Great Books program, the course is demanding and reading-intensive. Due to its reputation as a challenging and engaging class, the course is very popular among students. All of this results in varying degrees of self-selection into the course of students who prefer challenging courses or have a high "need for cognition," which research has found to be highly correlated to moral reasoning development (Cacioppo, Perry & Kao, 1984; King & Mayhew, 2002; Mayhew & King, 2008). Additionally, student orientation leaders are known to direct highly achieving and ambitious students to the course, adding to possible confounding of selectivity of students enrolled in the course and thus involved in the study

The course provides twelve credit hours per academic year, with three Philosophy and three Theology credits earned in the fall and spring semesters (and as such is weighted as a "double" course, representing 40% of the credit hours of a typical first year student's coursework) and serves as part of the university's core, liberal education curriculum. In its cornerstone course for first year students, students and faculty engage in tracing the roots of major philosophical and theological inquiries through close readings of primary texts, perennial texts. Among the explicit aims of the course is a critical and reflective reading of foundational texts from the disciplines of philosophy, theology, political science and ethics. Implicit goals of the course include an attendant

consideration of the ways that a student's values and moral sensibilities have been shaped by her culture and history. Heavily influenced by Heideggerian insights into historical-critical methodology, the course seeks to make students more consciously and critically aware of the evolution of contemporary notions of what is good, true, and valuable for individuals and for communities. Taking up ethical theory explicitly at points throughout the course, students engage in an evolving dialogue between ancient, medieval, modern, and contemporary thinkers about the moral aspects of personal, social and political orders. Textually centered discussions seek to encourage students to think critically and foundationally about the private and public natures of ethical and moral choices (for the common course syllabus, see Appendix S).

Faculty members who teach in the program assert its efficacy in advancing students' moral reasoning capacity in addition to their knowledge of foundational texts in the fields of philosophy and theology, but their evidence has been anecdotal. For this reason, assessment of the implicit goals of the course was desired. The data examined in this study represent a first attempt by program administrators and faculty to gain a large-scale assessment of these and other goals of the program.

3.5 DATA, DATA COLLECTION AND SAMPLE

3.5.A. Data

Experimental data involves a set of variables: a dependent variable which is the observed and measured response to some intervention or treatment, and an independent variable which consists of the intervention, presumed to be or not to be (in the null hypothesis) connected to the effect noted in the dependent variable. In this case, students' moral reasoning was the dependent variable while the course, understood in this

context as an educational intervention being “applied” to students, was the independent variable. Applying a null hypothesis, the study presumed that there would be no statistically significant change in the moral reasoning capacity of students who took a time one test (at the beginning of the course) to a time two test (at the end of the course), as measured against other samples and when controlling for potentially confounding factors, including age, maturation, gender, race, and political affiliation.

The data examined in this study were components of a 2012 program assessment of the year-long, freshman-level interdisciplinary course described above. A group of 16-20 faculty members who teach regularly in the program were consulted in the development of an assessment of the explicit goals of the course, which included facility with major themes of foundational texts in the history of Western philosophical and theological thought, and of implicit goals such as advancing ethical and moral reasoning. In order to capture data pertaining to the implicit aim of increased moral reasoning, faculty agreed to ask students in their sections of the course to complete the Neo-Kohlbergian DIT in a pre- and post-test fashion. Faculty also agreed to a program-wide assignment of an open-ended essay in which students were asked in the final weeks of the course to highlight class themes they felt had been particularly impactful to them and to their ability to address the central course question, “What Is the Best Way To Live?”

3.5.B. Data Collection

The DIT was administered in September, 2012, during the first few weeks of the fall semester and then again in April, 2013, during the final two weeks of the two semester course, seven months after the first test time. Faculty members were asked by the program director during the summer prior to the start of the academic year to allow

this assessment tool to be administered in their classes. The request was sent to faculty via email. Out of a total of 19 sections, 16 faculty members chose to participate in the DIT pre- and post-testing. Three faculty members either did not respond to emails about testing times and procedures or contacted the program director too late in the first semester to fully participate in the study. The measurement tool was described in some detail to faculty and 16 sections of the course were involved in this assessment. The tests were administered during class time, with all instructions read by the principle investigator in 14 of the 16 test instances. In two classes the test was administered by the classes' own faculty members.

Of the 452 students enrolled that year (AY 2012-13) in 19 sections of the liberal arts program used in this study, 385 students in 16 sections participated in the DIT survey. As recommended by the "Guide for DIT-2" (Bebeau & Thoma, 2003), students were given as much time as they needed and received no other direction beyond the instructions read from the DIT guide. Each student was assigned a 5-digit identification number and student names were erased from the tests for anonymity in the scoring process. Two copies of the list of identification numbers were kept by the researcher in a secured location and one copy of the list was kept in a secured location by the program director. The sets of tests were sent in two waves for scoring to the University of Alabama Center for the Study of Ethical Development. Due to class attrition, student absences or incomplete test forms (names were not included on 6 post-test answer sheets and thus could not be paired with a pre-test) 31 pre- and post-tests could not be paired and were thus invalidated. Of the 354 completed and paired pre- and post DIT tests from the sample, 29 additional tests were purged from the study based on the Center's own

well-researched battery of DIT reliability validity and reliability checks (Bebeau & Thoma, 2003). In total, 325 DIT pre- and post-tests made up the sample of the quantitative component of the study.

Additionally, students in all of the sections of the course were asked at the end of the course to write open ended essays discussing class themes that were particularly important to them and to the course's central question, "What is the best way to live?" However, in several sections of the course students were asked to write an essay addressing this question in the first week of class in September and then asked to write the final essay addressing the same question in relation to course content in the last two weeks of April, at the end of the course. Thus, essays from these classes have the quality of pre- and post-essays with respect to the course as an intervention. As such these essays illustrate some of the perceived impact of course content on student development and were thus examined via a qualitative analysis in the present study to complement the pre- and post-course quantitative assessment. This component of the study is discussed in detail in Chapter 4.

3.5.c. The Sample

The study began with of a convenience sample of 385 subjects who took the DIT as part of an assessment of a course in which they were enrolled as first year students and which serves as the educational intervention of the study. Of those original 385 students, the population of the quantitative component of the study consists of the 325 students who completed scorable pre- and post-DIT tests. Students participated in the DIT and the common written assignment as part of the course itself and as such, it was a captive sample. However, the sample used for the qualitative data section of this study

consisted of 49 students enrolled in just two of the sixteen classes involved in the study. In these classes, the essay assignment was assigned in the first week of class and then again at the end of the course and thus took the form of a pre- and post-treatment evaluation. Due to the fruitful nature of this type of format and its illustrative capacity, the students in these two classes became the subpopulation for the qualitative portion of the study. Of the 49 students who participated in the two classes in which pre- and post-course essays were assigned, only 46 completed both beginning and end of year assignments, due to class attrition. Thus, the 46 students in the qualitative subgroup represent 14% of the study sample. Limitations inherent in these sample selections are considered in a broader discussion of the study limitations at the end of Chapter 4.

3.6. THE QUANTITATIVE INSTRUMENT: THE DIT

3.6.A. History and Development of the DIT

Neo-Kohlbergian researchers James Rest and his colleagues at the University of Minnesota have contributed more than thirty years of research on the moral development of college students and young adults, primarily via their development and use of the DIT. Working with insights into Kohlberg's main tool of measuring moral reasoning, Rest developed the DIT, a paper and pencil test in which subjects read and responded to a series of narrative dilemmas, including the Heinz dilemma of the Kohlberg's Moral Judgment Interview (Rest, 1986; Thoma, 2002). The DIT utilizes prototypic statements and an accompanying set of complex decoy statements to prompt subjects to respond to various stage-related justifications for various resolutions. Thus, the DIT relies on a *recognition* model over a *production* model to explore subjects' attempts to resolve conflicts and sort out the complexities of dilemmas, a strategy that helps reduce

confounding effects of verbal and cognitive advantage. DIT subjects are asked to consider six moral dilemmas and then to rate and rank the relative importance of twelve statements related to a resolution of each dilemma. A series of scores are gathered from the rating and ranking tasks to produce a P-score, indicating the extent to which a subject used principled reasoning in working out the dilemmas (Rest, 1975, 1979, 1987, 1999; Rest, Narvaez, Bebeau & Thoma, 1999; Rest, Narvaez, Thoma, & Bebeau, 1999, 2000; Rest, Thoma, & Edwards, 1997).

Over two decades of DIT-based research offered the Minnesota group keen insights into the relative power of the P score. This research in turn helped the group reject other scores that failed to outperform it, including an early U score (Davison, 1977; Evens, 1995; Lawrence, 1987; Thoma, 1994, 2002; Thoma, Rest & Davison, 1991). Eventually, a more precise index was developed, known as the N2, which combined P scores with decreases or systematic rejections of lower stage reasoning (Lind, Hartmann & Wakenhut, 1985; Rest, Thoma, Narvaez, Bebeau, 1997). This combined power of two effects allowed researchers to identify when subjects are simultaneously utilizing complex moral reasoning and rejecting overly simplistic or inadequate reasoning, thus establishing a more discernable way of delineating what researchers mean by moral development generally. This new score also offered a robust mode of observing the influence of educational interventions on the moral reasoning capacity of students (Bebeau & Thoma, 1994; McNeel, 1994). As described in the review of the literature (see Chapter 2), the group's use of schema theory over hard stage paradigms helped highlight shifting distributions of stages as opposed to hard or discrete stages, thus locating the extent to which a subject *tends to use* higher or more complex levels of moral reasoning.

This allowed researchers to evaluate patterns of stage *orientations* on a continuous scale and to identify a subject's application of generic prior knowledge to the understanding of new knowledge as she parses out various aspects of new and more complex questions, attempts to fill in missing information, and pursues further relevant information toward an adequate resolution of a question or quandary (Rest, Narvaez, Bebeau, Thoma, 1999; Rest, Thoma, Narvaez, Bebeau, 1997; Thoma, 2002). The revised DIT is a capable tool to activate moral schemas so that an individual's *working conceptions* of moral principles can be evaluated. This development offers wonderful opportunities for researchers to examine moral development in the college experience.

3.6.B. The DIT in Detail

The present study utilized the current version of the DIT (the DIT2; for a full version, see Appendix A) which includes five dilemmas modeled on Lawrence Kohlberg's "Heinz dilemma," a moral quandary/dilemma central to Kohlberg's Moral Judgment Interview (MJI) widely used as a benchmarking tool in the early years of contemporary studies in moral development (Colby & Kohlberg, 1987a, 1987b; Rest, Narvaez, Bebeau, & Thoma, 1999). In the Kohlbergian dilemma subjects were asked to consider the actions of a man whose wife is dying of a type of cancer which could possibly be treated by a drug developed and sold by a local druggist. Heinz cannot afford the drug, the cost of which has been raised to 10 times its production value by the druggist, who will not sell the drug for less. Heinz breaks into the store and steals the drug. The updated DIT2 begins with a similar story, involving a man who is contemplating stealing food for his starving family from a wealthy man who is holding food supplies in a warehouse in order to sell the food at top value (Rest & Narvaez,

1998). Subjects are asked to read each story and then complete three tasks related to the dilemma.

First, participants are asked to consider and choose which action the protagonist should follow, to steal the food or not, or to indicate that they “can’t decide.” Interestingly, this first task is not involved in the eventual scoring of the DIT. It merely situates the issues of the dilemma within a horizon or context of action. Offering only short and simple options highlights that a choice or movement is necessary, triggering a demand for some sort of consideration. Second, subjects read a list of 12 statements that present issues related to resolving the dilemma, such as, “Does the rich man have any legal right to store food when other people are starving?” Subjects are asked to rate the significance of each item to the story on a 5 point Likert scale (rating each as having “great,” “much,” “some,” “little,” or “no” importance). Finally, subjects are asked to reconsider the 12 statements and choose four items they consider to be most important, ranking them as “most important,” second most important,” and so on. It is notable that the directions for rating and ranking the items are somewhat fragmentary. *To what* these items are significant or important is not made clear. Participants are left to draw their own conclusions about whether they are being asked about the importance of these items to themselves, to the characters in the story, to the resolution of the dilemma, or to society generally in rating and ranking tasks. This strategy allows participants to determine a full range of personal, social, or societal considerations involved in their choices.

The DIT’s prototypic statements were developed by Rest and his colleagues based on research into comments of hundreds of subjects through years of Kohlbergian MJI research, thus offering a highly nuanced representation of advanced and advancing moral

development (Rest et al., 1999). DIT researchers claim that the careful and thoughtful design of the DIT addresses many of the validity and reliability concerns of measures of this kind. For instance, extensive use of MJI research in the design of DIT statements counters a validity concern regarding the possibility of a highly deductive, Rawlsian bias of Kohlberg's earlier work. DIT statements allow other philosophical positions such as utilitarian and libertarian stances to be represented in the category of postconventional thinking. DIT researchers also claim that the DIT's recognition model (versus a production model) offers several benefits: first, recognition tasks resolve reliability concerns of interviewer and rater/scorer inconsistency; second, recognition tasks allow tacit understandings to be activated and reduce the degree to which verbal ability may confound developmental findings; third, recognition of items within rating and ranking tasks clarifies a horizon of macro- and micro- moral concerns for the subject's consideration and thus avoids validity issues associated with participant test interpretation. Lastly, the "fragment strategy" employed in the creation of DIT items defers to schema theory's insights into the centrality of recall and embedded memory in moral reasoning, strengthening the validity of the test (Narvaez, 1998).

3.6.C. Scoring the DIT: From P Scores to the N2 Index

The first versions of the DIT reported moral reasoning development in terms of a P score, originally designed to identify a subject's preference for prototypic statements based on Kohlbergian stages 5 and 6. As Rest and his colleagues developed working schemas (Personal Interest, Maintaining Norms, and Postconventional) in place of Kohlbergian stages, the P index came to indicate a subject's use of Postconventional modes of reasoning (see Chapter 2). By 1997 the Minnesota Group developed an even

more robust index for the DIT, the N2 score, reflecting the group's increased understanding of and confidence in schema theory over stage theory (Rest, Thoma, Narvaez & Bebeau, 1997). These researchers claim that the N2 index is a better fit for measuring moral schemas since it reveals tacit and general knowledge structures that are stored, invoked, activated, revisited and eventually chosen by subjects in moral reasoning (Rest, Narvaez, Bebeau, & Thoma, 1999). Additionally, this index highlights schema transition and consolidation processes, offering a nuanced construal of a subject's movement in and out of various schemas (Bebeau & Thoma, 2003; Rest et al., 1999).

P scores were used by James Rest and his colleagues for many years, despite their search for a more robust index that could outperform it. P scores are calculated via the subject's DIT ranking task only. Four points are scored if a subject assigns as "most important" a postconventional item from the 12 prototypic statements, three points are added to the P score if a postconventional item is ranked "second most important," two points are added to the P score for a postconventional item ranked third most important, and one point is added for a fourth ranked postconventional item. Thus, up to 10 points could be scored on each dilemma. Six dilemmas were used in the original version of the DIT to make a perfect P score of 60 points (the DIT2 uses five dilemmas with a total score of 50 points). The score was converted to a base 100 percentage, with scores ranging from 0-95% since not every dilemma included four postconventional items. In the original DIT, missing data resulted in the recalibration of scores on the total ranking points completed. In other words, if a subject failed to rank any item fourth on one dilemma, the score was recalculated based on total possible points of 59 rather than 60 (Rest et al., 1997). By 1997, the Minnesota Group discovered the work of German

researcher Georg Lind, who noted in DIT *ratings* a second indicator of development: a greater differentiation between high ratings of higher stage items and low ratings of lower stage items accompanied advances in development. Thus, the Group devised a new score which brought this additional component to bear on the DIT's measurement capability.

3.6.D. Calculating N2 scores

N2 scores combine a subject's use of postconventional thinking in the *ranking* task (based on the P score) with a second effect derived from a subject's systematic rejection within the *rating* task of more simplistic, lower stage thinking (Rest et al., 1997). The N2 index thus allows attention to be given to a subject's increased differentiation in moral reasoning rather than simply adverting to her advances in postconventionalism. To calculate the N2 score, two components are used. The first component, identified here as N2¹, is derived through an analysis of the subject's *rating* of each of the prototypic statements. The 12 statements read by subjects regarding the dilemma represent thinking in stages 2-6 (Stage 1 is omitted since the DIT's reading comprehension levels are aptly fitted to Stage 2-6 moral reasoning). The first component is a calculation of the subject's discrimination of stages 2 and 3 (which combined form the Minnesota Group's *Personal Interest Schema*) from stages 5 and 6 (the *Postconventional Schema*). In other words, a subject is not simply being scored on her preference for higher stage thinking or consolidated schema reasoning but for a clear discrimination of higher and lower stage thinking. The average of a subject's rating of stage 2 and stage 3 items is subtracted from the average rating of stage 5 and stage 6 items. This difference provides a discrimination measure which is then divided by the subject's standard of deviation of these four stages. Thus, the formula of the rating

component, $N2^1$, may be represented as: $N2^1 = \bar{x}\{\text{Stages } 5+6 - 2+3\}/\sigma$ (Stages 2+3+5+6).

The second component of the N2 score, identified here as $N2^2$, is derived via an analysis of the *ranking* task of the DIT, in which subjects select the four items from the twelve prototypic statements which they consider to be “most important.” Subjects identify or rank the first, second, third and fourth most important items, scored with four, three, two and one point, respectively, as in the original P score model. In this revised component, however, many of the scoring adjustments of the original P score are excluded and missing rankings are accounted for differently. If a subject omitted a ranking in the original DIT, the P score was recalculated based on the highest possible score of completed rankings. In the $N2^2$ no adjustment is made for a subject’s failure to complete all rankings. The omission of a rank is included in the calculation, understood in this version of the test as a failure to choose a Postconventional item. Thus, a total of ten points is possible for each of the five dilemmas, for a test total of 50 points. A subject’s score is converted to a percentage in the $N2^2$ component.

The DIT2 handles other omitted ranking and rating tasks differently as well. A subject’s failure to rank *any* items in one dilemma is adjusted for by using the total $N2^2$ score of the other four completed dilemmas. However, if the ranking task is incomplete on more than one dilemma, insufficient test-taking motivation is assumed and the entire protocol is invalidated. Similarly, if more than 4 items are not rated, the $N2^1$ is calculated based on the ratings of the other four dilemmas, but if 4 or more items are not rated on more than one dilemma, the test is invalidated. In other words, the DIT2 requires that the tasks of 4 dilemmas must be attempted with at least 9 items rated in each and a total of 14

of the 20 rankings completed. As Rest and his colleagues developed these new scoring procedures, they noted that drawing data from two DIT tasks might cause more frequent invalidation of tests resulting perhaps in smaller samples than the earlier version of the test. However, they felt confident that the new scoring model was robust enough to warrant the shift.

Finally, the two components of the N2 score, the N2¹ *rating* data and the N2² *ranking* data, are combined into one score by adding N2² to N2¹ weighted three times, based on research findings that *ratings* data have about 1/3 the standard deviation of the original *ranking* P scores. Thus weighting ratings data serves to equalize the rating and ranking data within the final N2 score. Rest and his colleagues submitted this N2 scoring to a battery of comparisons with a 1995 standardized sample of P scored tests, n=1,115, in order to standardize P-scored and N2-scored research and to demonstrate the new scoring model's sensitivity to a variety of validity measures. By the late 90s, the DIT2's N2 index was regularly outperforming P scores on a variety of important construct validity criteria that had long been posited by Rest and his colleagues as grounding for the DIT's strength as a measure of moral reasoning (Rest, Narvaez, Bebeau & Thoma, 1999a, 1999b).

3.6.E. DIT2 Validity and Reliability

It is important to note that the DIT's P and the DIT2's N2 scores are found to be approximately normally distributed. In a 1995 compilation of findings from a mega-sample (n=45,856) of DIT research culled between the years of 1989 and 1993, a mean distribution of 39.1 was found in a range from 0-91, with a standard deviation of 14.84, attesting to a normal distribution of P scores (Rest et al., 1999a). Thus, 66% of DIT

scores in the sample fall between 24.26 and 53.94 on a scale of 0-91, and 95% of DIT scores in the sample fall between 9.42 and 68.78. These results included data from DIT subjects representing a range of demographics and with a wide array of educational, socioeconomic and geographic characteristics. Results of the sample also highlighted the DIT's ability to capture postconventional reasoning, unlike much of Kohlberg's own MJI research. Current indices of the DIT2 have been standardized to these scales, with a range of 0-95 (Rest, 1987; Thoma & Rest, 1999).

Rest and his colleagues also operationalized seven validity criteria for assessing the construct validity of the DIT, along with test-retest and within-test reliability checks. The validity criteria include 1) differentiation of age and education levels, 2) longitudinal upward trends, 3) correlation to cognitive capacity measures, 4) sensitivity to moral educational interventions, 5) links to "prosocial" behavior and highly valued job performance, 6) links to political attitudes and choices, and 7) adequate reliability. Researchers have also demonstrated test-retest reliability and within-test consistency checks to safeguard against "garbage data" (such as answers that are selected to form graphic designs on answer sheets) (Rest et al., 1997, 1999a).

It is helpful to consider each validity criterion separately:

- 1) *differentiation of age and education levels*: Rest and his colleagues opined that tests of moral reasoning would likely show that graduate students in moral philosophy would score higher on the DIT than high school or junior high students. However, in order to determine if the test actually captured developmental differences, rather than educational opportunity differences or confounding variables like the socioeconomic advantages of those who attained

higher education, researchers used subsampling from very large composite samples to decrease the possibility of impactful confounding variables. Studies of these large composite samples showed that 30-50% of DIT variance is attributable to education level. Interestingly, Rest and his colleagues found evidence that after high school, subjects' stage 2 and 3 thinking decreases in proportion with stage 4 thinking in such a way that makes stage 4 thinking (Maintaining Norms Schema) redundant in N2 scores (see more about Stage 4 thinking and its significance in validity criterion #6 below). This explains why N2 scores are calibrated via increasing stage 5 and 6 thinking combined with decreasing stage 2 and 3 thinking, without regard for stage 4 thinking (Rest, 1999a).

2) *longitudinal upward trends*: DIT researchers utilize a 10-year longitudinal study of women and men from diverse educational, geographical and socioeconomic backgrounds to demonstrate that the use of postconventional thinking develops in a general upward trend. Moreover, Rest asserts that dozens of studies of college students attest to DIT gains as “one of the most dramatic longitudinal gains in college of any variable studied in college students” (Rest, Narvaez, Bebeau, & Thoma, 1999b, p. 310) with average effect sizes of .80 in liberal arts colleges and universities and similar effect sizes in the majority of college studies reported (Rest, Narvaez, Bebeau, & Thoma, 1999a).

3) *correlation to cognitive capacity measures*: the component of moral development that is ostensibly measured in the DIT is moral reasoning, which Rest and his colleagues asserted to be correlated to cognitive capacity. Using a modified version of Lind's Moral Comprehension Test (Lind, 1979, 2010, 2013;

Lind, Hartmann, & Wakenhut, 1985), which assesses a subject's ability to correctly identify and recall the gist of a moral argument, the Minnesota group verified a significant ($p < .001$) correlation between moral comprehension and P and N2 scores. This significant sample correlation coefficient (moral comprehension with P is $r = .67$ and with N2, $r = .69$) of DIT scores to moral comprehension, recall and reconstruction of moral arguments and articulations provides an important piece of evidence of the DIT's validity.

4) *sensitivity to moral educational interventions*: Researchers assumed that if moral reasoning was indeed a part of development connected to cognitive growth, it would respond to educational programming and intervention. By the late 90s, DIT research included over 60 published studies that examined educational interventions via the DIT, including a 1985 meta-analysis (Schlaefli, Rest, & Thoma) of 55 studies that showed moderate gains (.41) following educational interventions that lasted longer than three weeks, compared to small gains (.09) for shorter treatments. Moreover, older subjects (college and older adults) showed greater change, adverting to researchers' assertions that educational interventions work best in advancing higher stage thinking after adolescence (Rest et al., 1999a). Mathew Mayhew's extensive body of work in the past ten years examines a plethora of educational interventions' impact on moral reasoning using the DIT, finding significant effects in a variety of intervention types (Mayhew, 2012; Mayhew & Fernandez, 2007; Mayhew & King, 2008; Mayhew, Seifert, & Pascarella, 2010; Mayhew, Wolniak, & Pascarella, 2008).

5) *links to “prosocial” behavior and highly valued job performance:*

Measuring the link between moral reasoning and moral behavior has been a long and hotly disputed point in Kohlbergian and neo-Kohlbergian research. Rest and his colleagues readily concede the point that thinking through moral dilemmas remains in the world of the hypothetical and cannot uncover a subject’s real moral action or behavior. Kohlbergian theory and research was criticized vociferously as over-intellectualizing moral development. Many continue to maintain against neo-Kohlbergian work that *thinking about* moral choices cannot be linked to actually *making* moral choices. Over the years, however, researchers have continued to mine this area of study, noting the important distinctions made by James Rest between various components of moral development. As such, Rest and his colleagues have delimited the DIT as a measure not of moral character (the ability to persist in moral tasks and behavior) but of moral judgment, by which a person judges which action is most justifiable (Rest et al., 1999a). More recently, members of the Minnesota Group have joined other researchers in assessing professional decision making, performance ratings of healthcare, accounting, management and education professionals (Bebeau, 1994, 2001, 2002; Bebeau & Brabeck, 1987; Bebeau & Thoma, 1994, 1999; Rest & Narvaez, 1994). DIT research also suggests that P and N2 scores are significantly correlated at the same levels to “prosocial” behaviors such as community involvement and civic responsibility as measured by service to one’s community. Rest acknowledges that this validity criterion offers the weakest association, accounting for only 5-20% of the variance of behavior measures (Rest et al., 1997, 1999a).

6) *links to political attitudes and choices*: Since the DIT proposes to measure what neo-Kohlbergian theory understands as macromorality, or the way that people relate to others within social horizons, moral reasoning is conceived as illuminative of political attitudes and decision making. Hence, the Minnesota Group assumed that P and N2 scores would map on to particular civil libertarian attitudes regarding issues including free speech, religious toleration, human rights, etc. The group found that in studies from the 1970s through the late 1990s, P scores remained relatively consistent, notable especially in light of shifts in American political views throughout that time period. They also noted that P and N2 scores highly correlate with political attitude in the $r = .40-.60$ range and in some cases account for over 60% of variance in attitudes towards highly controversial issues (including abortion, women's rights, free speech, etc.). Interestingly, in this validity criterion, the P score offers an advantage over the N2 score, as Rest and his colleagues found in their work with the development of a Law and Order scale early on in DIT research. Wanting to examine closely the shift to principled morality, researchers specified in this scale the shift from "prioritizing social order, unquestioned deference to authorities, and rejection of deviance to the prioritizing of individual welfare, questioning of authority, and tolerance of deviance" (Rest et al., 1997, p. 503). Using this scale, Rest and his colleagues demonstrated that P scores correlated more closely to attitudes of political toleration and awareness over and against law and order preferences than N2 scores. Additionally, P scores offered the possibility of isolating and thereby highlighting this unique shift from the law and order orientation of the

Maintaining Norms Schema (Stage 4) to Postconventional thinking (Stages 5 and 6) by simply subtracting Stage 4 thinking from P scores. This “P-Stage 4” index underlines the final shift out of the Maintaining Norms Schema into the Postconventional Schema by calculating P at the expense of Stage 4 thinking, as opposed to the N2’s calculation using the exclusion of Stage 2 and 3 thinking (Bebeau & Thoma, 2003; Narvaez, Getz, Rest & Thoma, 1999; Rest et al., 1997). As regards political orientation, the DIT asks students to identify themselves as very liberal, liberal, neither liberal or conservative, conservative, or very conservative. Subjects who classify themselves as conservative or very conservative report higher Maintaining Norms scores, while those who classify themselves as liberal or very liberal have higher P scores, though research has shown that moral judgment and political orientation do not reduce to each other (Emler, Palmer-Canton, & James, 1998; Maeda, Thoma & Bebeau, 2009; Walker, 2002). Thus, moral judgment is viewed by the Minnesota Group as a “co-contributor to political choice rather than a proxy for political orientation” (Bebeau & Thoma, 2003, p. 40). However, the difference has been found to become significant at the sophomore level of college and since the present study examines the DIT scores of freshmen, the study did not focus on this aspect of DIT comparisons.

7) *adequate reliability*: As to the internal reliability of the DIT, researchers have put forward composite samples of DIT studies to demonstrate its Cronbach’s alpha to be in the high .70s and low .80s, with the N2 index coming in consistently in the low .80s. Cronbach’s alpha is a measure of a test’s internal

consistency and is widely used to evaluate whether items in a test actually measure one thing, in this case, moral reasoning. Educational practice looks for a standard of .80 (variance of .36), which is achieved consistently in the DIT. Research has shown that ideal ranges of Cronbach's alpha fall in the .70 to .95 range, though scores higher than .90 often suggest redundancies in test design (Tavakol & Dennick, 2011).

3.7. CONCLUSION

As evidenced in this study's review of the literature of moral development, quantitative data gleaned from the DIT is extremely useful in observing particular aspects of moral growth. However, what precipitates moral development, especially in the context of college participation and its impact on advancement in moral reasoning remains largely shrouded. Decades of moral development research attempting to ascertain the efficacy of educational interventions have primarily utilized quantitative scales such as the DIT and the MJT. More recently, researchers like Mathew Mayhew have combined quantitative measures to identify and scale other confounding variables that impact moral growth within educational environments. Mixed methods approaches in this area are not common, likely due to the array of problems that plague mixed methods research (including problems of representation, legitimation and integration) and researchers' subsequent tendencies to favor one analytic format over another (Onwuegbuzie & Johnson, 2006). The present study opted for a mixed methods approach to addressing the efficacy of educational interventions, suggesting that a coherent, holistic picture of moral development and growth would be aided by student's own perceptions of what moved them forward intellectually and personally within the span an

educational experience. If the objective of moral development research is to discover and encapsulate aspects of moral advancement, qualitative data's ability to extract unique and illustrative themes and motifs from subjects' reflections should be considered alongside the insights that quantitative measures provide. In the next chapter, the qualitative data of this mixed method will be presented.

Chapter 4:

Qualitative/Rubric Methodology

4.1 INTRODUCTION: RATIONALE FOR QUALITATIVE ANALYSIS OF WRITING

While Defining Issues Test (DIT) data offer a unique snapshot of student moral development, it is important to consider the challenges of determining appropriate and sufficient measures for moral or character outcomes. Moral development is neither simply defined nor easily operationalized for measurement and outcome assessment, and a variety of proxy measures are often used in quantitative research of moral development for the sake of convenience and positivistic science aims (Dalton, Russell & Kline, 2004; Strange, 2004). Qualitative and quantitative research in college student moral development often suffer from limitations, including small sample sizes, single-institution cases and reliance on potentially biased student self-reporting. Though the DIT and similar measurement tools are designed to control for confounding factors, neo-Kohlbergian research's insistence that the DIT is a measure of only one, narrowly construed component of moral development reminds us of the difficulty in assessing an aspect of personal flourishing that is so broadly understood and holistic. Moreover, common notions of advanced moral development anticipate the ability to follow through on moral understandings in moral tasks (Rest's *moral character* component) and this developmental component eludes most research tools. In order to capture as clear a picture of moral development as possible, therefore, a combination of measures is warranted, as noted in Chapter 3. A mixed methods approach allows for:

the strategic and purposeful combination of both quantitative and qualitative data collection and analysis...assum[ing] that the epistemological and methodological advantages of each can work in concert to corroborate or more robustly support the findings, or to reveal complementary or even contradictory outcomes (Saldaña, 2011, p. 10).

Combining qualitative analysis of writing and DIT data provides a means to focus a descriptive and analytic eye on specific attributes of moral development as well as on the impact of educational interventions on that development. The present study triangulated quantitative and qualitative data in an effort to observe in these particular cases what, if anything, was happening in a course of study that was thought to be impactful to student development. In looking at student writing from a select group of students from the study sample, the study engaged in a data source triangulation model (Stake, 1995) to uncover aspects of student development that are evidenced (or not) in DIT scores. Triangulation of DIT data and subsequent groupings of students based on DIT scores with student writing also allowed a consideration of what a measured moral development gain or loss “looks like” from the perspective of students’ own experiences of reflecting on important topics, texts and themes. In this model different types of data are linked together, inviting both cross-validation of particular dimensions of a research question and further interpretations of the meaning of research findings (Plano & Creswell, 2008; Stake, 1995).

4.2 THE RESEARCH QUESTION

Thus the study utilized a mixed methods approach to examine students’ moral reasoning development, including a qualitative analysis of essays written by students

about their own experiences of the course and course content. The purpose of the second, qualitative component of the study was two-fold and addressed the project's second and third research questions: 1) how do students perceive the moral dimensions of the course? and 2) how is student perception of these moral dimensions connected to gains in moral reasoning development? In concert with quantitative, DIT data, student writing was analyzed to address these questions and in so doing, to provide a fuller picture of students' moral development.

4.3 THE RESEARCH DESIGN

As noted in Chapter 3, an initial analysis of the DIT results of the total study sample were completed in the first stage of analysis. The second, qualitative component of the analysis focused on a convenience sample subgroup of 46 students from 2 classes (see Chapter 3's description of the samples used in the study). An evaluation of this subgroup's DIT scores allowed subjects to be divided into five categories based on scoring attributes:

- 1) *Low to Low*: students whose pre- and post-test scores were low compared to national trends, with limited or no change (losses or gains ≤ 12 points [approximately 1 standard deviation]);
- 2) *High to High*: students whose pre- and post-tests were high compared with national trends, with limited or no change (losses or gains ≤ 12 points [approximately 1 standard deviation]);
- 3) *Gainers*: students who posted significant gains in post-tests, regardless of position on continuum (gains ≥ 12 points [approximately 1 standard deviation]);

- 4) *Decliners*: students who posted significant declines in post-tests, regardless of position on continuum (losses ≥ 12 points [approximately 1 standard deviation]);
- 5) *Constant Average*: students who posted pre- and post-test scores commensurate with national averages and whose scores remained steady (gains or losses ≤ 6 points [approximately .5 standard deviation]).

Essay analysis focused on the written work of subjects who fell into the first four categories, working with subgroups of not fewer than three students in each subgroup. Subjects who remained constant and average in DIT scores were omitted from analysis. The study identified three or more students in each group from the subsample for qualitative investigation to avoid focusing on outliers. Analysis of three or four student essays from each group 1) revealed language patterns and articulation of concepts, attitudes and behaviors that offer points of convergence or divergence with the set of a priori moral development proxies identified in the first rubric (see Section 4.6 for rubric details); and 2) identified common course topics, content, texts, discussions, themes, etc., identified by students as impactful to their own development within and across DIT groupings. This analysis was implemented via two sets of a priori rubric items designed by the principle investigator to highlight moral development proxy items and course-related items that connect student experiences of course content with aspects of moral development.

The two rubrics used in the study were developed via a thorough review of the literature of moral development and a consideration of the course content and program aims (see rubric details in Sections 4.5 and 4.6 of this chapter). Writing from subjects in all four groups was double-coded along both rubrics. Rubric analysis utilized a typology

development strategy to uncover the homogeneity within and heterogeneity between subgroupings of subjects categorized along DIT scoring patterns (Caracelli & Greene, 1993). In other words, analysis was undertaken first to identify what student moral development “looks like” within the various groups, ie. what sort of thematic language and concept patterns emerged within each group that were connected to moral development proxies; second, student self-reporting of what they considered to be particularly impactful in the course and in the course content were analyzed to identify connections between course-related items and moral development markers. Motifs and concepts were identified via three levels of coding (open/pattern, axial and selective coding, described in the Analysis Section below). The analysis provided a picture of what might be particularly impactful to the moral advancement of students *in different stages within moral development* and how students perceived the connections between course content and their own moral development.

4.4 DATA AND DATA COLLECTION

Students in almost all sections of the course intervention (for course description and details, see Chapter 3) received a fairly open-ended essay assignment at the end of the year in which they reflected on the various texts and themes of the course to answer a central question, “What Is the Best Way to Live?” Faculty members in the program agreed that gathering essays for course assessment would complement other assessment tools and most agreed to submit selections of essays to the program director. In the academic year of the study (2012), 19 sections of the course were offered. Out of those 19 sections, students from 16 sections of the course participated in the DIT pre- and post-testing, with 325 students with complete and scorable DITs included in the study sample

(see Chapter 3, section 3.5.C. for a detailed description of the sample). Of the 16 classes, 2 sections posed the assigned essay in a pre- and post- course fashion, asking students to reflect on the question, “What Is the Best Way to Live?” The essay was assigned during the first week of class in September and was then repeated during the final week of the spring semester. These essays were chosen for qualitative analysis to complement DIT data since they offered a unique view of students’ own sense of their progress from the beginning of the two-semester course to its end.

Student essays from these two classes represented a small but not insignificant percentage of larger sample of students who participated in the DIT study (n = 15 student essays, 5% of study sample). A selected group of essays was examined via a deductive heuristic to uncover 1) evidence of moral reasoning or moral sensitivity that illustrate particular DIT findings or illuminate DIT gains; and 2) connections between student self-reported perceptions of the course and course content and their own moral development. A protocol for assessing various components of moral development within student writing included analysis via two rubrics: the first included a series of student attitude, belief and behavior outcome items that correlate with moral development research, while the second was comprised of course-related items that anticipate and organize connections between the course and students’ expressions of their experiences in the course. Student essays were double coded throughout this process for analysis via both rubrics.

4.4.A. Subsample Essay Selection Process

The subsample of the qualitative component of the present study originally consisted of 49 students from two classes in which students were required to write essays

at the beginning of the program and at its end. Thus, these essays were assigned at the same time that students were taking the DIT pre- and post-tests (early September and late April). In both classes, the pre-course essay prompt was the same: *“In 3 pages, answer the question, ‘What Is the Best Way to Live?’”* As such, the essays were construed in this investigation as pre- and post-course reflections on what a student considers valuable and worthwhile in constructing a life lived well, also one of the central themes of the course. The post-course essay prompts were also the same for both classes: *“In 3 or more pages and in light of the course, how would you now answer the question from the first essay”* Thus, the second prompt did not explicitly ask students to *refer to* course texts or themes, though students would likely anticipate that as an implicit part of the end-of-year assignment. The subsample included 49 students enrolled in these two classes (25 in one class, 24 in the other). Attrition, unmatched pre- and post-DIT tests, or incomplete essay assignments resulted in a final subsample of 46 subjects (94% of the subsample and 14% of the total sample). Using SPSS, frequencies and pre- and post-test differences were used to compare subsample Pre-test N2 scores, Post-test N2 scores, N2 change scores, etc. with those found in the total sample (for details, see Chapter 5). The categories used to identify notable groups in the total sample were revised slightly in order to choose subject essays for qualitative analysis. A full discussion of the rationale and procedure of this adjustment is presented in Chapter 5.

Six months prior to subsample essay selection, essays of all students in the two classes were matched with DIT identification numbers (using the master list), labeled with those ID numbers and names were blackened out by the primary researcher. Since the primary researcher was also the instructor of one of the classes, concealing authorship

of the essays was necessary to ensure as randomized a selection process as possible. The final essay selection was more than 10 months after the class ended and thus authorship was no longer identifiable to the primary researcher. Lists of Gainers, Decliners, High-to-High Scorers and Low Gainers were then generated from this subsample using SPSS analytics, which produced a list of pre- and post-test N2 scores and N2 Change Scores for all 46 students in the subsample. Using SPSS case sorting, students within this subsample were identified as belonging to the various groups outlined for analytical purposes.

4.4.A.1 Selection of Gainer Essays

Nineteen identified Gainers (N2 Change Score ≥ 12 points [or gain of 1 SD]) represented 38% of the subsample, a higher percentage than the 27% of Gainers found in the total sample, though the subsample Gainers' mean gain score of 18.69 was lower than the total sample's mean gain score of 20.33. Five essays were selected randomly from the list of 19 Gainers. One of the essays belonged to the student whose gain score was the maximum gain score of 38.57, and though this gain score is certainly notable, the subject's pre-test score was also extremely low (10.95). Thus, that essay was excluded on the basis of its outlier position. The four remaining randomly selected essays belonged to students whose mean N2 Change score was 20.27, considered acceptable since the group's mean fell within one point of the mean change score of the total sample. Once analysis was complete, ID numbers were checked against DIT lists and it was revealed that the four Gainer essays were written by female students and were evenly divided between the two classes involved in the subsample.

4.4.A.2 Selection of Decliner Essays

Decliner essays were selected from a list of subjects who fit the parameters of the Decliner category (N2 Change Score ≤ -6 points [or loss of .5 SD]). Only 5 decliners were identified, representing 10% of the subsample with a mean loss of -9.42 (SD=3.02). The essay which represented the maximum N2 Change score loss (-13.55) was excluded in favor of the remaining four whose mean loss was -8.39 in order to avoid the inclusion of an outlier. Both the subsample decliners and the decliners whose essays were selected from the subsample posted mean losses that were lower than the total sample mean loss of -11.79 (SD=6.35). After analysis was complete, comparison of decliner essays against DIT information revealed that three of the decliners were female students, one male, with the decliners once again evenly divided between the two classes (2 per class).

4.4.A.3 Selection of High to High Essays

Five students of the subsample were identified in the High to High group (pre- and post-test N2 Scores ≥ 54 points), representing 10% of the subsample, as compared with 12% of the total sample. The group's mean Pre-test N2 Score was 59.37 (SD=3.17) with post-test N2 Score mean of 62.25 (SD=6.45) were consistent with the total sample's pre-test mean of 60.27 (SD=4.51) and post-test mean of 62.44 (SD=6.28). Of the five students in the group, one was selected because the subject posted a slight decline from pre- to post-testing (though not enough to move below the cut point). All of the other subjects in the group gained in N2 post-tests. Three more subjects were chosen randomly from the remaining four for a total of four essays from High to High scorers. Of the four students in this group, 3 were male and 1 was female and all were from the second class (the class in which the primary researcher was the instructor).

4.4.A.4 Selection of Low to Low Essays and New Gainer Group

Because Low-to Low scorers were so rare in the subsample (n=2), only one of whom completed both pre- and post-course essays, this group proved too small for adequate analysis. Thus, the category was omitted from the qualitative component of the study. Descriptive statistical findings about the group remain helpful from a programmatic point of view and to aid faculty in understanding the range of students they might encounter in their courses within the program, but the group was not an adequately represented for the purposes of this investigation.

Due to the inadequate representation of the Low to Low scoring group in the subsample, the primary researcher expanded the examination of subjects who posted gains during the year, but whose pre-tests scores fell in the total sample's lowest quartile. This second category within the Gainers group ("Low-Gainers") was created in order to examine the characteristics of students whose pre-test scores were low relative to national norms and to scores within the sample.

To create this new Low Gainer group, an SPSS quartile frequency analysis was used to sort the original 19 Gainers' scores into pre- and post-test quartiles. This process subdivided the Gainer group into Gainers who gained from very low scores, with pre-test scores in the first quartile (henceforth labeled Low Gainers), versus those who had gained from the second or third quartiles into higher quartiles (Gainers). As it turned out, the four subjects randomly selected for essay analysis in the initial Gainer group each had pre-test scores that fell in the second or third quartile and post-test scores in the third or fourth quartile (see Table 6.2) so they retained the label, Gainers. The quartile analysis of N2 Pre-test scores and N2 Post-test scores of the 19 subsample Gainers yielded a distinct group of 7 Low Gainers (36% of subsample Gainers and 15% of the subsample) whose

pre-test scores were calculated in the lowest quartile but whose post-test scores moved them into higher post-test quartiles.

Thus, the new group, Low Gainers, included students who began with very low scores relative to the total sample (the lowest fourth of the sample) and gained enough in N2 scores over the course of the year to move out of the lowest quartile of the sample by the post-test. Though the N2 Pre- and Post-test scores of this Low Gainer group were on average quite a bit lower than the average of the Gainer group (N2 Pre-test means 22.59 versus 30.12), their gains across the sample were higher than other Gainers (N2 Change means 24.79 versus 20.33). Still, since the study seeks to consider the program's impact on students across the full range of moral reasoning development, there is a clear benefit in incorporating this group of students who come in at significantly lower points in that range in the study.

From the group of seven Low Gainers in the subsample (14% of the subsample), three Low Gainers (6% of the subsample) were randomly selected for essay analysis, including 2 females and 1 male. All Low Gainers were from one class, not taught by the primary researcher of the study.

4.5 DESIGN OF RUBRICS AND RUBRIC ITEMS

4.5.A. Rubric Item Development

Creating a qualitative research tool capable of capturing aspects of moral reasoning via a close examination of student writing involved the challenging prospect of operationalizing outcomes related to moral development and flourishing. A review of the literature of college student moral development research revealed an array of sources from which to draw attitudes and behaviors that signal moral development. Researchers

have utilized a variety of findings from DIT studies and other measures of moral, ethical and character development in identifying proxies for moral development and moral reasoning. Moreover, researchers such as Pascarella (1997, 2005a, 2005b) and Mayhew (2004, 2012) have explored the intersection of moral reasoning measurements and moral, character and ethical development gains reported in large student surveys. These types of research projects provided rubric elements which were tailored to the study's aim of connecting moral reasoning data with this specific program's course content as well as its explicit and implicit goals. Finally, an overview of moral and ethical development rubrics and benchmarking from a variety of higher education foundations and associations offered a final consideration of plausibly helpful and reliable rubric items for the analytic tool.

4.5.B. Rubric items from DIT studies

Extensive use of the DIT2 in research examining a variety of variables that might impact or be impacted by moral reasoning development was very instructive for developing a writing assessment tool. Mayhew (2002, 2004a, 2007, 2008a, 2008b, 2010a, 2012a, 2012b) has examined the impact of pedagogical practices, classroom environments, course characteristics and the course taking patterns and behaviors in a decade of active moral development research with the DIT2. He has also explored a variety of outcomes hypothesized to be impacted by moral reasoning, including social justice attitudes and behaviors, political orientations, rejection of moral exclusion, etc. Mayhew notes that discomfort or dissonance about addressing complex moral or ethical issues may also be a hallmark of moral development, insofar as “individual social and cognitive development is a function of disequilibrium and the extent to which an

individual can reconcile one's own perspective with those of another" (2007, p.59). Thus, this analysis looked to student writing for levels of intellectual humility and an acknowledgment of not fully knowing what a just or moral life might be like, akin to the ideal posed by Socrates (in Plato's *Apology*, a required text in the course considered in the study) that a wise person knows that she does not know all that is to be known.

4.5.c Rubric Items from meta-analyses and survey benchmarking

A number of researchers have conducted meta-analyses of moral reasoning studies to consider various characteristics of students, college environments, and educational interventions related to moral development. The scholarly conversation around moral development research now includes work that combines or nests DIT studies within other, large surveys such as the Wabash National Study of Liberal Arts Education (WNS) (Mayhew, Seifert & Pascarella, 2012; Mayhew, Seifert, Pascarella, Laird & Blaich, 2012; Mayhew, Wolniak & Pascarella, 2008), the National Survey of Student Engagement (NSSE) and the College Student Experience Questionnaire (CSEQ) to detect patterns of moral attitude and behavior (Kuh & Umbach, 2004; Pascarella, Seifert & Blaich, 2011). Using benchmarks from these surveys to analyze proxy variables found in these large databases, researchers have uncovered a variety of additional elements to explore and assess moral development among college students. While limitations related to large surveys' dependence on self-reporting remind us that subjects tend to overestimate their own developmental gains (Bowman & Seifert, 2011; Pike, 1995, 1996), leaders in the field of moral development research and learning assessment have much to contribute from these substantial research projects. Likewise, national organizations, foundations and research clearinghouses like the AAC&U, the Teagle

Foundation and the Dalton Institute on College Student Values' Character Clearinghouse offer insightful rubric items from wide ranging research in the field of moral development.

4.5.D Piloting of Rubrics

In order to increase the reliability of the essay analysis, the primary researcher piloted the rubric application with a researcher experienced in qualitative analysis and coding. Five anonymous essays were randomly selected from the total group of sample essays to use in the piloting of the rubric analyses. The two researchers reviewed the essays separately, applying items from both rubrics, and then rating each rubric item. Researchers used a rating system to consider the extent to which the essay author demonstrated proficiency and/or proclivity for each of the moral reasoning proxy and course-related items within the rubrics. Researchers met on three occasions: first, to review rubric items and definitions, second, to establish parameters of the rating system, and third, to compare essay analysis, ratings and themes of piloted essays. The primary researcher utilized "Hyperresearch" software to label and rate phrases, sentences and sections of texts to which rubric items were applied. The second researcher hand-coded and rated the essays.

During the pilot researchers determined that a scale of 1-3 allowed analysts to determine *to what extent* students exhibited proficiency in or proclivity (the tendency or inclination to choose a particular attitude, behavior, or value) for moral reasoning rubric items, and *to what extent* students were engaged in and/or by course-related items. The rating system also examined whether students demonstrated egocentric, emerging or integrated attitudes and understandings of moral proxy items. In other words, did student

writing exhibit an *egocentric* or primarily self-referential position with respect to the moral proxy item in question, or did the essay demonstrate an *emerging* or more fully *integrated* proclivity toward the item? Piloting helped clarify words, phrases, themes and questions within student writing that foregrounded the analysis process. For instance, while in one piloted essay a student acknowledged the value of showing interest the good of others (associated with “Prosocial Attitude” in the moral reasoning rubric), the student only reflected on how that might be applied to her/himself, not how s/he might engage in that value with others. The rating system was not used to validate DIT scoring but to detect patterns in student’s own reflections on what makes a “good” life and in their self-reporting about the course’s impact on their attitudes and values. The rating system also sharpened of the researcher’s attention to nuances within the various themes and motifs in the essays.

The pilot essay analysis included an extensive discussion about the best way to approach to the rating system. Researchers considered neo-Kohlbergian characterizations of Personal Interest, Maintaining Norms and Postconventional schemes along with other developmental scales in a thorough vetting of the rubric items’ meanings and interpretations. It is important to note that the rating system was used not to validate DIT findings but as a strategy to maximize open coding reliability. A rating of “1” was assigned to a writing that demonstrated a low proficiency, low proclivity or egocentric stance with respect to a moral reasoning item. A rating of “2” was assigned to writing that displayed a moderate and/or emerging proficiency with respect to the moral rubric items and a moderate or emerging engagement with the course-related items. A rating of “3” was assigned to a writing that consistently demonstrated and seemingly integrated

proficiency in moral reasoning rubric items or a consistent engagement with and understanding of course-related items. For instance, in rating the moral development rubric item, “Prosocial behavior,” the two coders who piloted the rubric agreed on a rating of “3” for an essay that thoughtfully and elaborately demonstrated how volunteering with physically and intellectually challenged adults and children during the year had impacted her thinking about the best way to live. For this same rubric item, the coders rated another essay “1” in light of that author’s highly abstracted suggestion that helping others would be a good idea, seemingly for others to do. For the most part, disagreement or uncertainty between the coders addressed the rating of “2” and thus a more regularized notion of what counted as “emerging” or moderate proficiency. Coders agreed that a student mentioning the importance of a rubric item was not enough, but that a “2” would be warranted if the student adverted to some apparent movement toward the value or ideal.

In the course of piloting the rubric items, several items consistently overlapped in almost every instance. For instance, the moral development rubric item, “Civic engagement attitude,” which is operationalized as the “expressed desire to make a difference in communal contexts” and “Prosocial attitudes,” operationalized as “shows interest in the good of others without anticipation of personal reward” were used to tag the same aspects of essays at the same rating in just about every essay, and thus these items were collapsed in the final rubric analysis. The operationalized definition of one moral development item, “Diversity,” proved to be vague and open to several interpretations by the researchers and was thus revised via a review of the relevant literature as “the ability to articulate a position from the point of view of someone outside

one's 'insider' group." With respect to course-related items, "Use of texts in reflection," which referred to students' use of course texts in writing, appeared so consistently and at such high levels on post-course essays that it seemed to lose relevance and was thus omitted from the final course-related item rubric. Another course-related item, "References to text and course activities," proved to be more helpful since this item sought a demonstration of student *understanding* of concepts within texts and course activities rather than a simple mention of them. Finally, it was determined that within course-related rubric analysis, a rating of "0" would be helpful to indicate *no engagement with or mention* of a course-related item. Thus the range of ratings used in the course-related rubric was adjusted to 0-3, as opposed to the 1-3 rating system used on the moral development rubric. Since the ratings of these rubrics were not combined, this adjustment had no further implications.

By the final meeting, the two raters achieved a 93% agreement rate on moral development rubric items and a 90% agreement rate on course-related rubric items. Thus, levels of rubric validity and rater reliability were considered sufficient to move forward with rubric analyses. The final rubrics are presented here in Chapter 4.

4.6. RUBRICS

A two-pronged, rubric-based evaluation of student essays began with a first rubric that identified evidence of attitudes, understandings, behaviors or beliefs connected to postconventional moral reasoning. Analysis via open, descriptive, axial and selective coding of data served to illustrate and highlight a variety of proxies that connected to moral developmental gains (Saldaña, 2011). This first rubric was based on a review of moral reasoning literature, with particular attention paid to results of meta-analyses and

large national surveys, as well as an overview of selected pilot essays. Table 4.1 outlines this rubric's 15 moral development proxy items for coding along with operationalized definitions and scholarly references for each item. While moral development rubric items serve as *a priori* codes, an analysis of essays allowed additional items that were central and uniquely illustrative of moral reasoning to emerge.

First, essays were open-coded to identify statements that connected with rubric items and other items that emerge as possibly connected to student moral growth. Via an analysis of language frequency and depth of engagement, coded essay selections were rated as exhibiting no, low, moderate or high proclivity or engagement relative to rubric items consistent with coding protocols developed in the piloting of the rubrics with a second rater (See complete details of the rubric pilot in Section 4.5.D). A second cycle of axial coding established thematic clusters and coding categories found in the essays. Finally, selective coding was applied to axial codes to ascertain specific themes and categories that are associated with moral development.

In a second rubric analysis, student writing was analyzed via *in vivo* coding of students' own reporting of their perceptions of the moral content and dimensions of the course. The list of 12 course-related items comprising this second rubric design was derived from both the literature review and a thorough consideration of the course description as well as stated course and program aims (gleaned from the university catalogue, department web site and the program's own self-description). In some cases rubric items refer to specific course content (eg. "Aristotelian ethics") and in other cases to course objectives regarding ethical benchmarks (eg. "Theory and real world

connection”). Table 4.2 presents the 12 items of this analytic rubric in addition to operationalized definitions and scholarly source and course description identification.

In vivo coding analysis of these essays and a subsequent cycle of axial coding of selected highlights from essays offered direct access to student expression of the connection between their own experiences of the course and course aims. “HyperResearch” software was utilized in all of the qualitative analysis processes and was especially helpful in this part of the writing analysis. In this analytic process, specific phrases, words and sentences from student essays that relate to these various course-related items were gathered and then clustered to produce a picture of which practices, themes, texts and ideas of the course were particularly fruitful for students in their own development. Selective coding was construed based on a consideration of how faculty may best use insights from student essays to highlight areas of strength of the course and uniquely impactful themes, texts, discussions, assignments and ideas of the course.

Table 4.1 Writing Rubric Items, Moral development proxies for descriptive, axial and selective coding

Moral		
Development Item	Operationalized definition	Scholarly source
Prosocial attitudes	Shows interest in the good of others without anticipation of personal reward	Cooper, Liddell, Davis & Pasquesi, 2012; Gibbs, 2013; Gilligan, 1982; Kuh & Umbach, 2004; Rest et al., 1999b, 2000
Prosocial behaviors	Adverts to social action intended for the benefit of others without anticipation of personal reward	Cooper, Liddell, Davis & Pasquesi, 2012; Gibbs, 2013; Gilligan, 1982; Rest et al., 1999b, 2000
Civic engagement attitude	Expressed desire to make a difference in communal contexts	AAC&U Rubric, 2011; Astin & Antonio, 2004; Colby, Ehrlich, Beaumont & Stephens, 2003; Cooper, Liddell, Davis & Pasquesi, 2012; Gibbs, 2013; Reed, 2008
Diversity	Acknowledge and comprehend a variety of perspectives on issues and multiple worldviews	AAC&U Rubric, 2011; Colby & Kohlberg, 1987; Gibbs, 2013; King & Mayhew, 2002; Kuh & Umbach, 2004; Pascarella, Seifert, & Blaich, 2010; Seifert, Goodman, King, & Baxter Magolda, 2010
Personal and social responsibility	Sees oneself as having agency in the welfare of others and the larger community	Colby & Kohlberg, 1987; Cooper, Liddell, Davis & Pasquesi, 2012; Gibbs, 2013; Hersh & Schneider, C, 2005; Hoffman, 2005; Kuh & Umbach, 2004; Mayhew & King, 2008
Critical thinking	Demonstrates capacity to synthesize ideas, images, skills to address problems in innovative ways	AAC&U Rubric, 2011; Kuh & Umbach, 2004; Pascarella, Seifert, & Blaich, 2010; Seifert, Goodman, King, & Baxter Magolda, 2010
Global thinking	Attempts to acquire a comprehensive exploration of issues and problems of the global community	AAC&U Rubric, 2011; Colby & Kohlberg, 1987; Cooper, Liddell, Davis & Pasquesi, 2012; Hoffman, 2005; Rest et al., 1999b, 2000
Socratic ideal	Recognizes and acknowledges the limitations of one's own understandings, contexts, and experiences	
Cognitive complexity	Demonstrates a capacity to exclude simplistic models of problem	Cacioppo et al., 1984; Colby & Kohlberg, 1981; Mayhew, 2012; Mayhew & King, 2008; Mayhew, Wolniak & Pascarella, 2008;

	solving	Narvaez & Bock, 2002; Pascarella, Seifert, & Blaich, 2010; Rawls, 2001; Rest et al., 2000
Social complexity	Rejects exclusionary thinking as regards social interactions and communal living	Colby & Kohlberg, 1981, 1987; Cooper, Liddell, Davis & Pasquesi, 2012; Gibbs, 2013; Hoffman, 2005; King & Mayhew, 2002; Mayhew & King, 2008; Pascarella, Seifert, & Blaich, 2010; Rawls, 2001
Principled thinking	Articulates universalizable principles and a desire to live according to widely held principles	Colby & Kohlberg, 1987; Gibbs, 2013; Mayhew & King, 2008; Pascarella, Seifert, & Blaich, 2010; Rest et al., 1999a, 1999b, 2000
Context comprehension	Acknowledges the limiting factors of context and signals a grasp of one's own limited context	AAC&U Rubric, 2011; Cooper, Liddell, Davis & Pasquesi, 2012; Mayhew & King, 2008; Narvaez & Bock, 2002
Ethical self-awareness	Attempts to express core beliefs and an understanding of their origins	AAC&U Rubric, 2011; Blasi, 1980; Colby, Ehrlich, Beaumont & Stephens, 2003; Gibbs, 2013; Hoffman, 2005; Rest et al., 1999b, 2000
Ethical evaluation	States an ethical position and provide an articulate summation and defense of it	AAC&U Rubric, 2011; Colby & Erlich, 2000; Colby, Ehrlich, Beaumont & Stephens, 2003; Kuh & Umbach, 2004; Narvaez, 2008; Rest et al., 1999a, 1999b, 2000; Seifert, Goodman, King, & Baxter Magolda, 2010
Moral behavior	Expresses change in behaviors toward self and others	Blasi, 1980; Gibbs, 2013; Hoffman, 2005; King & Mayhew, 2002; Pascarella, 1997; Pascarella, Seifert, & Blaich, 2010
Expansive religious attitude	Sees religious understandings as cause for expanding circles of care and concern rather than exclusion	Astin & Antonio, 2004; Burwell, 1992; Getz, 1984; Maeda, Thoma & Bebeau, 2009; Murk & Addleman, 1992; Narvaez, Getz, Rest & Thoma, 1999

Table 4.2 Writing Rubric Items, Course –related items matched with in vivo codes

Course –related item	Operationalized definition	Scholarly source
References to texts & course activities	Demonstrates an understanding of ethical or moral concepts as articulated in texts, movies, discussions, etc. of the course	Halliday & Frantis, 2006
Use of texts in reflection	Utilizes texts to frame the project of on-going reflection and reflective thinking	Halliday & Frantis, 2006; King & Mayhew, 2002; Mayhew & King, 2008; Narvaez, 2008
Theory & real world connection	Demonstrates increasing ability to connect theories of justice and real-world issues	Bruess & Pearson, 2000; Cacioppo et al., 1984; Grunwald & Mayhew, 20008; Kuh & Umbach, 2004; Maeda, Thoma & Bebeau, 2009; King & Mayhew, 2002; Mayhew & King, 2008
Course content as interruption	Shows evidence of ethical and moral ideas that have caused disequilibrium or have problematized earlier assumptions	Maeda, Thoma & Bebeau, 2009; King & Mayhew, 2002; Mayhew & King, 2008; Rest, 1986, 1999a, 1999b
Ethical-Political theory	Articulates newly formed or formulated understanding of political theory regarding common good	Reed, 2008; <i>Perspectives in Western Culture Syllabus and Course Description</i> , Braman, 2013
Ethical theory	Articulates newly formed or formulated understandings of ethical or moral theory	Colby & Erlich, 2000; Halliday & Frantis, 2006; Narvaez, 2008; Reed, 2008
Aristotelian ethics	Shows an understanding of situation-based virtue ethics and demonstrates a nuanced view of the role of practical wisdom in assessing ethical choices	Linstrum, 2009; Narvaez, 2008; <i>Perspectives in Western Culture Syllabus and Course Description</i> , Braman, 2013
Religious/Natural Law ethics	Shows an understanding of underlying foundations of morals and ethics and notices how these are shaped by inherent ends or ultimate purposes	Narvaez, Getz, Rest & Thoma, 1999; <i>Perspectives in Western Culture Syllabus and Course Description</i> , Braman, 2013
Kantian ethics	Demonstrates a recognition that principles of goodness and justice are needed to clarify these notions for a community	Linstrum, 2009; Narvaez, 2008; <i>Perspectives in Western Culture Syllabus and Course Description</i> , Braman, 2013

Common Good tradition	Articulates the need for nuanced understandings of social order and the role of institutions; recognizes competing interests of community and individuals	Reed, 2008; <i>Perspectives in Western Culture</i> Syllabus and Course Description, Braman, 2013
Nietzschean/Existentialist ethics	Recognizes the dignity of individuals and the primacy of spiritual values over and against the practical needs of society	<i>Perspectives in Western Culture</i> Syllabus and Course Description, Braman, 2013

4.7 DATA ANALYSIS - RUBRIC APPLICATION

The qualitative component of this study examined student writing from four subgroups of students, determined by DIT scores (see Section 4.3 for details of subgroups). As such, this aspect of the study utilized what Stake (1995) refers to as a *collective case study*, in which two or more cases are analyzed in order to learn more about something else. This study used student writing from these groups to understand more about student development reflected in DIT scores and how educational experiences in college impact that development. Stake maintains that while observation of one or even several cases “is a poor basis for generalization” it can offer a means to “refinement of understanding” (1995, p. 7). Triangulating data sources, such as this study’s yoking of quantitative data and case studies, aids a capacity to interpret what is happening beyond or behind data points. Here, analysis of student writing illuminated common and disparate themes from students’ educational experiences associated with moral development by identifying what students themselves considered to be important personal insights, questions and experiences from the course.

From each of the 15 essays selected for rubric analyses, two sets of observations were captured: first, a set of clustered themes that emerged from coding related to the first rubric’s moral development proxy items, and second, observations pertaining to course themes, topics, texts, ideas, and classroom experiences reported by students to be impactful. The analysis process identified patterns *within* groups (Low Gainers, High to High scorers, Gainers, and Decliners) regarding moral development items and course-related items. Analysis compared patterns *between and across* groups. Both

commonalities and differences illustrated what the various DIT scorers “look like” with respect to moral development proxy items and impactful aspects of the course.

Two questions were posed in light of the first set of coding of moral development proxy items: 1a) which proxy items most notably stand out among High to High Scorers and Gainers and what are the similarities among them? and, 2a) which proxy items are conspicuously missing in Low Gainers and Decliners and what is similar about these missing components? Two additional questions were posed regarding the second set of observations of impacting course-related items: 1b) which classroom or course-related experiences are considered important and are actively reported by High to High Scorers and Gainers and which items are underreported? and, 2b) which classroom or course-related items are important to Low Gainers and Decliners and which course experiences are underreported by these groups?

A final interpretive stage of analysis explored the sorts of course-related experiences (encounters with texts, movies, discussions in and outside the classroom, etc.) that were connected to various moral development proxies for students at different stages of moral reasoning development. This interpretive step pivoted back to DIT data from the larger sample of 325 study participants. SPSS analytics were used to determine the percentages of students who fall into the parameters of the four groups. Quartile analysis offered general moral development benchmarks for the types of students who enroll in this program and revealed a number of helpful observations about students within and across groups. Findings are suggestive for programmatic pedagogical strategies aimed at enhancing and expanding the moral development of students at different stages of that development.

4.8 LIMITATIONS OF THE STUDY

Several limitations of the study may have impacted the extent to which findings of the study may be generalized regarding moral reasoning development. Literature in the field of moral development includes a number of important critiques of measures like the DIT which attempt to concretize and statistically assess what many consider to be an abstract construction of cognitive and affective functions (Elm & Weber, 1994; Gilligan, 1982; Maxwell, 2010; Emler, Palmer-Canton & St. James, 1998). It also includes myriad critiques of the theoretical underpinnings of neo-Kohlbergian research and its normative claims. While several decades of neo-Kohlbergian research using the DIT has provided greater precision and clarity along these lines, the concerns remind moral development researchers to be wary of grand claims and overreaching conclusions. Recent neo-Kohlbergian research which engages new insights in neurobiology and psychology suggest promising avenues to resolving aspects of moral development that have previously been difficult to identify and assess (Gray, Young & Waytz, 2012; Narvaez 2001; Narvaez & Bock, 2010; Shu, Gino & Bazerman, 2011; Young & Saxe, 2011). This type of research is in a relatively nascent stage, however, and it suffers its own set of constraints.

With respect to the qualitative analysis of students' essays, research demands a careful consideration of the extent to which student's self-reporting of their own developmental gains correspond to actual gains. Recent research, including a longitudinal sample of over 3,000 college students found students to be fairly inaccurate in assessing their own gains in cognition and personal development (Bowman & Seifert, 2011). Additionally, Pike (1995, 1996) suggests that while student *self-knowledge* is mostly

adequately self-reported, it seems that self-reporting of *gains* in development is not. It seems that knowing one's present capacities is less challenging than fairly assessing what gains have been made. Across gender, race, academic achievement and institutional type Bowman and Seifert (2011) found startlingly low correspondence between self-reports of gains and longitudinal gains on the same construct, noting that subjects, especially young adults and students, tend to report socially desirable responses which threaten the validity of many studies that depend on self-reporting.

In addition to these research-wide limitations, there are several important limitations of the study design that should be highlighted, first in terms of the sample of subjects involved in the assessment and second, in terms of aspects of the qualitative methods:

1. The study itself lacks an adequate control group. Since the data were construed as part of a course evaluation, there was no control group established. Moreover, several aspects of the course make finding a control group unusually difficult, including its two-semester design and its restriction to first year students. A possible consideration for a future study along these same lines might include using a control group of first year students who wanted to take the class but were not able to register due to limited course space. This design would resolve some of the confounding effects of student self-selection into a course that is known to be challenging.

2. The students in the study sample were not randomly assigned to the course. Their self-selection into a course known for its rigor and limited access adds a set of sample selection biases that demand careful attention as regards findings and conclusions about those findings. Students who would choose this type of course would

likely have high “need for cognition,” an attribute which is found to significantly correlate to high DIT scores (Cacioppo, Petty & Feng Kao, 1984; King & Mayhew, 2002).

3. It should be noted that the primary researcher was not only an instructor in the program but was also the instructor of one of the two classes from which the student essays were selected. In light of this, concerns of researcher bias threaten the validity of the study’s findings on two fronts. The primary researcher had prior knowledge that the DIT would be administered and understood its components quite well. Though the contents of the DIT were not discussed in the class, the primary researcher might have unconsciously “taught to the test” in a way that might not have occurred in other sections of the course. Also, the primary researcher knew that the qualitative measure (the essays) would also be used in subsequent course assessment, though during that year there was no plan to connect the essays with the DIT. A second area of concern relates to the researcher’s knowledge of the students whose essays comprise 24 of the 46 essays available for the qualitative subsamples. Knowledge of student work during a year-long course offers challenges to unbiased coding of those essays and thus necessitated additional safeguards against possible preconceptions. To address these limitations, two methods were employed: first, the director of the course program, who had access to the list of identification numbers and corresponding student names assigned identifying numbers to essays and deleted names from essays to veil student authorship; second, a second rater/coder was selected to participate in a pilot of the rubric coding protocols for both the moral development item and course-related item rubrics. Once a set of themes, topics and important course-related items were noted, clustering those themes in the

second round and analyzing them in a third round of coding was done by the study's principle researcher. This second rater design provided a "second set of eyes" to identify impactful class topics, texts and experiences emerging from students' essay reflections. Thus, while researcher bias posed a potential challenge to the validity of this study, safeguards such as concealing student names on essays and employing a second rater in the coding phase of analysis mitigated these limitations.

4. Since the number of student essays used for this analysis and the number of classes they represent from the whole sample was relatively small, it is possible that the students in the two classes chosen for essay analysis were not representative of the group as a whole. The effects of class environment, pedagogical strategies, class dynamics, etc. may have significantly confound effects and thus skewed the qualitative findings. However, the strength of pre- and post-course design of the essays from these two sections of the course, taught by two different instructors, assures that analysis offered a unique illumination of various aspects of moral development and its relation to purposeful educational content and course design. The study's use of a collective case study model, furthermore, illustrates that the study does not intend to represent the moral development of all college students or to verify DIT findings, but rather to gain a clearer and more refined understanding of moral development.

The limitations inherent in researching moral reasoning and moral development were daunting to be sure. As noted in the literature review, limitations range from detailed sifting through confounding effects and apprehending components involved in moral development, to a more comprehensive hermeneutic of suspicion regarding the very notions of morality and of the human capacity for moral reasoning. Still, college

impact theory asserts that something unique in the college experience affects students' moral development, particularly in the first and second years (King & Mayhew, 2002; King, 2009; Mayhew & King, 2008; Pascarella, & Terenzini, 2005; Rest, 1999). To what extent this development is (or is not) articulated or understood and what aspects of educational intervention play a role in it necessitates further study in the area of moral development and what spurs its growth.

Chapter 5:

Analysis of Quantitative Data and Findings

5.1 INTRODUCTION

This study focused on two sets of data in an attempt to examine facets of moral reasoning development in students in this liberal arts program. Analysis of the study sample DIT scores offered an overview of the range of student moral reasoning development that faculty may encounter in this program, or in other courses like this, and essay analyses provided a fuller account of 1) what students' moral reasoning "looks like" in their own reflections on living a "good life" and 2) which aspects of coursework may have been particularly impactful in those reflections. The study did not seek to use essays to verify DIT scores but rather to examine what sorts of ethical and moral issues are commonly found among high and low scorers as well as among subjects who posted substantial gains or losses in DIT scores. Similarly, the study also sought to identify which aspects of this course were reported to be impactful or important by students in thinking about what makes a good life. Two rubrics were used to analyze the essays. A first rubric of moral reasoning proxy items was developed from a thorough review of the literature and a second rubric was designed with specific course-related items that are programmatically intended to raise moral and ethical questions. Chapter 5 presents the procedures and findings of the quantitative components of the study while Chapters 6 and 7 provide overviews of the procedures and findings of the qualitative analyses. The central questions of the study will be addressed as far as the findings allow; additional issues discovered in the analyses and further questions raised by the data are presented in the final chapter.

5.2 DIT SAMPLE AND DIT ANALYSIS PROCEDURE

As explained in details in Chapter 3, 325 subjects completed fully scorable pre- and post-course DIT tests (see Chapter 3 for a full discussion of the DIT and its scoring). This sample included 171 male students and 153 female students, with one subject not identifying gender. Scores from both pre- and post-course DITs were received in separate files from the University of Alabama's Center for the Study of Ethical Development. In order to create and examine N2 change scores, the files were merged in SPSS (IBM's "Statistical Package for the Social Sciences," a statistical analysis program) using ID numbers to match cases. N2 Pre-test scores were subtracted from N2 Post scores to create a new variable, "N2 Change," in order to identify change in subjects' moral reasoning scores over the time period of the study. Descriptive statistical analyses and paired sample t-tests were run on the N2 Score pre-test and post-test variables and on the N2 Change variable using a confidence interval of .95. These analyses garnered mean scores, measures of standard deviation, standard error of the mean, quartile descriptors and correlation information.

An internal reliability analysis of the study's DITs found a Cronbach's alpha of .59 for the pre-test and .62 on post-test variables, calculated via an SPSS reliability scale analysis of 5 variables that represent the N2 scores within each of the 5 stories. These figures are lower than typical DIT reliability rates which are typically found to be in the high .70s to low .80s. However, Bebeau and Thoma (2003) point out that the Cronbach's alpha will be lower when a full range of educational levels is not represented in a study. Since all of the subjects in the present study were in the same grade level, the Cronbach's alpha found in this case was acceptable.

5.3 DIT FINDINGS

5.3.A Sample N2 Mean Scores and National Norms

Three variables were used to capture a “picture” of the range of student moral reasoning development one might typically find among students in this program: subjects’ pre-course N2 scores, post-course N2 scores and N2 Change scores. Overall, students in this study scored much higher than national norms in both pre- and post-tests, with a pre-test mean of 40.75 at a standard deviation of 13.52 and post-test mean of 45.94 with standard deviation 12.20. Table 5.1 depicts the descriptive statistics of the pre- and post-test N2 scores of the sample. These scores may be compared with researched

Table 5.1 Study sample N2 Pre-and Post-test scores (N = 325)

	Minimum score	Maximum score	Mean score	Std. Error	Std. Dev.
N2 Score_Pre	1.73	71.22	40.75	.749	13.52
N2 Score_Post	14.07	75.59	45.94	.677	12.20

national norms for college freshmen which report a mean N2 score of 31.05 at a standard deviation of 14.42. It is helpful to situate the scores found in this study against national norms in order to understand possible implications of higher or lower scores and dramatic change scores among students in the sample. Indeed, the pre-test scores of this study sample more closely mirror national figures for Master’s degree students, who achieve mean scores of 40.56 with a standard deviation of 15.06. A comparison of sample N2 scores and national norms is depicted in Table 5.2. The present study’s sample post-test

scores come in even higher at 45.94 with a standard deviation of 12.20, rivaling nationally normed scores of students in graduate professional degree programs. To keep

Table 5.2 Comparison of N2 Scores and National Norms

	Ntl. Norm Freshman	Ntl. Norm Senior	Ntl. Norm MA Degree	Ntl. Norm Prof. Degree	Sample Pre-Test	Sample Post-test
N	2,096	2,441	853	1,582	325	325
N2 Mean Score	31.05	36.85	40.56	44.97	40.75	45.94
Standard Deviation	(14.42)	(15.53)	(15.06)	(14.87)	(13.52)	(12.20)

these relatively high scores in perspective, and to describe the sort of student who tends to self-select into this program, it is interesting to compare these scores to those found in a similar study of first year students in a business ethics program at the same university. In that study (Sullivan, 2011), first year students from the university's school of management who on average have posted the highest SAT/ACT scores coming into the university, logged pre-test N2 scores of 35.37 (with a standard deviation of 13.42) and post-test N2 scores of 40.16 (15.43), starting out in between the normal scores of college juniors and seniors and ending their first semester just shy of Master's degree level-normed N2 scores. Subjects of the present study sample coming from across the university and thus including students from the Colleges of Arts and Sciences, Nursing, Education and Management, posted significantly higher pre-test N2 scores. It is impossible to compare post-test scores between these study samples since the business ethics program consisted of only one term while the present study sample applied post-tests after a two-term program. Still, to understand the range of scores of this present sample and of students who opt to enroll in a rigorous liberal arts core program, it is

important to highlight that the sample's first year students posted higher scores in *pre-tests* than students at the same university scored *after* a full term ethics course.

One explanation for this difference within the same university with a presumably similar population could be explained in the breakdown of gender within the samples. The business ethics program study sample consisted of 71% male students (n=190) and 29% female students (n=78), while the present study was weighted more evenly at 53% male students (n=171) and 47% female students (n=153 [one student not identifying as either male or female]). Since female students' scores were higher in pre- and post-tests in both studies, it makes sense that the higher percentage of female students represented in the present study resulted in higher overall scores. However, female students in this study posted pre-test N2 scores with a mean of 42.24 (SD=12.35), higher than their female business ethics counterparts in Sullivan's 2011 study in which female students posted pre-test N2 mean scores of 38.68 (SD=14.14). In both studies, though female students posted significantly higher pre- and post-test N2 scores than their male counterparts, N2 change scores between males and females were not statistically significantly different in either study. See Table 5.3 for these comparisons. As discussed in the review of the research literature (see Chapter 2), several meta-analyses of DIT studies show gender to be a weak predictor of DIT performance, with educational attainment found in Thoma's 1986 meta-analysis of over 6,000 subjects to be 250 times stronger than gender in accounting for variance of moral reasoning. But gender disparity in moral reasoning development is more complicated as educational levels increase. Both Thoma (1986) and Bebeau and Thoma (2003) found gender differences to be less

impactful in younger subjects, with women's moral reasoning advantage over men widening at higher education levels. Boston College's admissions' strategy of pursuing a

Table 5.3 Comparison of Pre-test N2 Scores among same-university freshman women and men in Business Ethics course (Portico) and Perspectives course

N	N2 pre-test	mean score	Standard Deviation
Portico Study females	78	36.68	14.14
Portico Study males	190	34.01	12.90
Perspectives Study females	155	42.24	12.35
Perspectives Study males	177	39.34	14.39
National norms female freshmen	1,271	34.02	13.54
National norms male freshmen	808	29.66	14.07

(Note: Portico scores from Sullivan, 2011)

relatively gender-balanced population results in its female students boasting higher GPA and SAT scores than their male classmates. While the neo-Kohlbergian paradigm affirms an association of moral reasoning with cognitive-structural development, research with the DIT consistently finds that moral reasoning is not reducible to cognitive or intellectual capacity. Female students' higher DIT scores in this and similar samples may be connected to a range of variables including particular intellectual advantages or their advanced moral reasoning capacity. The consistency of higher female scoring among Boston College first year students raises important questions about the cognitive motivation, intellectual ability and moral sensitivity (a separate component of moral development according to neo-Kohlbergian theory thought to be more advanced in female subjects; see Chapter 2 for a fuller discussion) of the female students who select into the program but is unfortunately beyond the scope of the present study's research questions.

5.3.B Sample N2 Change Scores

As stated above, pre- and post-test results were merged using matched cases and a new variable, “N2 Change,” was created via a descriptive statistical computation of the variables “N2Score_post” minus “N2Score_Pre.” The new variable “N2 Change” identifies change in subjects’ moral reasoning scores over the time period of the study. Table 5.4 illustrates the results of subjects’ scores changes. Overall, subjects in the sample posted a mean gain of 5.19 (SD=12.35) over their first year, a gain which is closer to the sort of gains Thoma (1986, 2003) finds in college students over their whole college career.

Table 5.4 Descriptive Statistics of N2 Change

	N	Minimum	Maximum	Mean	Std.Deviation
N2_Change	325	-37.71	51.79	5.1871	12.346

A paired sample t-test of pre- and post-test N2 scores with a confidence level of 95% showed that student N2 score gains from pre- to post-test were statistically significant, with a t result of 7.57 (2-tailed significance). The comparison of pre- and post-test means with standard deviations pooled resulted in a significant correlation of .543 (significance $\geq .001$), as depicted in Table 5.5. A t-test of paired differences revealed that statistically significant gains in moral reasoning were made during the period from the pre-tests administered in September to the post-tests administered in late April of the year of the study. Thus, the general answer to the first research question, “Does the moral reasoning capacity of first year college students increase in a Great Books course with an implicit goal of promoting moral development?” is answered

affirmatively and the null hypothesis is rejected. Though moral reasoning development is expected on the basis of maturation alone, the sort of gains demonstrated in this sample is statistically significant. The scores of this sample are notable from both the perspective of the gains made within the sample and the mean scores achieved by the group as compared to national norms.

Table 5.5 T-Test Statistics, Correlations and Paired Differences

Paired Sample t-test statistics	Mean	N	Std. Dev.	St. Error Mean	Correlation	Sig
Pair 1						
N2 Score_Pre	40.75	325	13.52	.75		
N2 Score_Post	45.94	325	12.20	.678	.543	.000

Paired Differences	Mean	N	Std. Dev.	St. Error Mean	t	df	Sig.
Pair 1							
N2 Score_Pre-N2 Score_Post	5.19	325	12.35	.685	7.574	324	.000

5.4 GROUPINGS OF DIT SCORES

Quantitative data from the DIT and qualitative data from student essays were triangulated for the purposes of establishing groups for investigation. As such, groupings of DIT scores were identified based on mean N2 scores, mean N2 gains and losses and standard deviation calculations. These distinguishable groups highlight for instructors and program designers the range of student moral reasoning development that might be found in a group of first year college students encountered in a program like the one at the center of the study. The groups were:

- 1) *Low Gainers*: students whose pre- and post-test scores are low compared to national trends, with limited or no change (losses or gains ≤ 12 points [approximately 1 standard deviation]);
- 2) *High to High*: students whose pre- and post-tests are high compared with national trends, with limited or no change (losses or gains ≤ 12 points [approximately 1 standard deviation]);
- 3) *Gainers*: students who post significant gains in post-tests, regardless of position on continuum (gains ≥ 12 points [approximately 1 standard deviation]);
- 4) *Decliners*: students who post significant declines in post-tests, regardless of position on continuum (losses ≥ 12 points [approximately 1 standard deviation]);
- 5) *Constant Average*: students who post pre- and post-test scores commensurate with national averages and whose scores remain steady (gains or losses ≤ 6 points [approximately .5 standard deviation]).

Essay analysis focused on the first four categories unless the fifth, subjects who scores remained constant with little or no change, was found to be of particular interest in the DIT analysis. It turned out that the categories, High to High, Gainers and Decliners were the most fruitful and populous groups, with the Low to Low scorers difficult to find in large numbers. The category of Constant scorers did not prove to be interesting for the study since scores in this range simply tended to move only slightly and in no particular pattern. Additionally, the parameters of the Decliner category warranted adjusting since N2 change loss of 1 full standard deviation was uncommon. However there were a significant number of subjects whose scores declined from pre- to post-test, a fact interesting in and of itself in light of assumptions about moral reasoning development as

an upwardly directed developmental model impacted by age and education. Thus the parameter for identifying Decliners was adjusted to include N2 score changes of -6 points, or .5 a standard deviation, in order to avoid focusing on outliers in the qualitative component of the study. Essays written by students in this category represent a range of N2 change scores of -7.03 to -11.76. Thus this set captures the range of decliners in the subsample, the mean of which is -9.42 (SD=3.02).

Similarly, those whose pre- *and* post-test N2 scores were low compared to national numbers and remained low, the Low to Low category, were uncommon within the subsample for essay selection. Even after adjusting the parameter of the LL group to students whose scores came in below the national norms and whose posted gains or losses of .5 standard deviation or more kept them below the national average, only one student from the subsample was identified in this group for the essay analysis component of the study.

5.4.A DIT Groups from DIT analysis

In the final analysis, DIT groupings were identified using the following criteria:

- 1) *High to High*: students whose pre- and post-tests are high compared with national trends, with limited change (losses or gains ≤ 12 points [approximately 1 standard deviation]) (n = 40, 12%);
- 2) *Gainers*: students with pre-test scores in any quartile and who post significant gains in post-tests (gains ≥ 12 points [approximately 1 standard deviation]) (n = 88, 27%);

3) *Decliners*: students with pre-test scores in any quartile who post significant declines in post-tests, regardless of position on continuum (losses ≥ 6 points [approximately .5 standard deviation]) (n = 57, 17%);

4) *Low to Low*: students whose pre- and post-test scores started and ended low compared to national trends, with limited or no change (losses or gains ≤ 6 points [approximately .5 standard deviation]) (n = 17, 5% [only 2 in subsample])

5.4.A.1 DIT Gainers

Using an SPSS variable filter (N2 Change ≥ 12), 88 Gainers were identified from the study sample, representing 27% of the total sample. The group posted a mean gain in N2 Change scores of 20.33, with a standard deviation of 7.87, gaining as much as 51.79 points in the maximum instance in the time between the pre and post-tests. Table 5.6 highlights the case of Gainers from the sample group. These gainers cut across

Table 5.6 Gainers: Descriptive Statistics

	N/total	% of Group	Minimum	Maximum	Mean	Std. Dev.
Sample Gainers N2 Change	88/325	27%	12.26	51.79	20.33	7.87
Subsample Gainers N2 Change	19/49	38%	12.26	38.57	18.69	6.98

the spectrum of high and low scorers, with 54% of gainers coming from the lowest quartile (N2_Pre < 31.35) of pre-test scores, 25% coming from the second quartile (N2_Pre ≥ 31.35 and ≤ 41.45), 15% from the third quartile (N2_Pre ≥ 41.45 and ≤ 51.26), and 6% from the fourth quartile (N2_Pre > 51.26). Gainers represented 27% of the total study sample and 38% of the subsample (the two classes from which essays for

qualitative analysis were drawn). Within the subsample, the gains posted by this group showed a mean of 18.69 with a standard deviation of 6.98. The four subjects selected from this group for the essay analysis component posted gain scores within 4 points of the mean in order to glean a representative view of this group.

5.4.A.2 DIT Decliners

A surprising number of students posted negative changes in N2 scores from the beginning of the course to its end after two semesters. A full 17% of the total sample, 57 subjects, saw a decline 6 points or more (roughly .5 the standard deviation of the post-test N2 scores and N2 change scores which both had standard deviations of 12 rounded to the nearest whole number). As displayed in Table 5.7, these decliners saw a mean N2 change of -11.79 (SD=6.35), with a startling maximum decline of -37.71 in at least one case. Of the 57 decliners of the sample, 43 came from the top and third quartile of pre-test N2 scores, with a pre-test mean score of 48.70 (SD=10.82) mean. That is, 75% of decliners in N2 change scores came from the top half of pre-test scorers, well above national and sample norms. Only 5 of the decliners came from the bottom quartile of pre-test scores.

Table 5.7 Decliners: Descriptive Statistics

	N/total	% of Group	Minimum	Maximum	Mean	Std. Dev.
Sample Decliners N2 Change	57/325	17%	-6.0	-37.71	-11.79	6.35
Subsample Decliners N2 Change	5/49	10%	-13.55	-7.03	-9.42	3.02

Six decliners were identified in the subsample, enabling a random selection of four essays for the qualitative analysis. The N2 change scores of these four students fell within a standard deviation of the mean of the subsample group. Essays in this group present a

range of N2 change scores of -7.03 to -11.76 and thus are representative of the range of decliners in the subsample, the mean of which is -9.42 (SD=3.02), as well as the range of decliners in the total sample with mean change -11.79 (SD=6.35).

5.4.A.3 DIT High to High Scorers

High to High scorers represent the group of subjects whose pre-test N2 scores are a full standard deviation above the mean scores of the sample and at least two times the standard deviation above the national norms for college students generally. In this sample, the N2 mean score in pre-tests was 40.75. Adding one standard deviation of the pre-tests (13 points) to the pre-test mean, a cut point of 54 points was used to delineate scores for this High to High category on pre-tests. In order to avoid disadvantaging high scoring students, the post-test score cut point was kept at the same number, 54, since that threshold is far above researched expectations of college undergraduates' N2 scores. An initial criterion for remaining in the High to High category was the requirement that subjects not post *losses* of more than 6 points (.5 standard deviation), but this additional requirement was found to be too restrictive and was omitted. Only one student in the total sample actually posted a large decline (-13.55 points) from a very high score in the top quartile (N2_Pre score = 60.54). This case was so extreme that it was considered an outlier and though it fell within the parameters of the High to High group within the subsample, it was thus not selected for the subsample essay analysis.

The sample included 58 subjects whose pre-test N2 scores were above the cut point of 54, totaling 18% of the sample. By the post-tests, 40 students, representing 12% of the sample, remained above the cut point of 54 and thus in this High to High category. As stated above, when the additional criterion of not losing more than 6 points was added

to the cut point demarcation, only 29 (9%) students remain in this high range. Therefore, this criterion was dropped to loosen the restrictiveness of the category. The mean scores of the High to High group came in at a pre-test mean of 60.27 (SD=4.51) and a post-test mean of 62.44 (SD=6.28), as shown in Table 5.8. These scores are far above national norms for graduate students, matching those found in research examining the moral

Table 5.8 High to High Students in Sample (n=40/325, 12%) and Subsample (n=5/49, 10%)

	Minimum	Maximum	Mean	Std. Dev.
Sample HH N2 score pre	54.11	71.22	60.27	4.51
Sample HH N2 score post	54.03	75.59	62.44	6.28
Sample HH N2 Change	-9.82	16.57	2.17	6.70
Subsample HH N2 score pre	54.88	63.40	59.37	3.17
Subsample HH N2 score post	55.29	70.59	62.25	6.45
Subsample HH N2 Change	-4.77	10.05	2.88	5.45

reasoning levels of PhD/EdD students and Moral Philosophy graduate students. The subsample of two classes from which essays were drawn for qualitative analysis included 5 students in this High to High category, representing only 10% of that subsample. Essays from four of these students were selected for the qualitative analysis component of the study.

5.4.A.4 DIT Low to Low Scorers

Low to Low scorers were not difficult to identify within DIT scores, but their inclusion in the *qualitative* component of the study proved to be daunting. Low to Low scorers were identified as students whose pre-test N2 scores were a full standard deviation below the mean of the sample pre-tests, or below 27.24 (mean of N2 Pre 40.75

[SD=13.51]). This cut point was also several points below the national mean N2 score for college freshman of 31.05. The post-test N2 scores of Low to Low students was identified as scores remaining below the mean of the sample pre-tests (40.75). Only 17 students in the total sample, or 5% of the sample, fell into this category, with only 2 (4%) from the subsample. Table 5.9 gives an overview of some descriptive statistics of this group. Of the two students within the subsample who qualified in the Low to Low group, only one completed both pre- and post-course essays, and so the group was omitted from the qualitative component of the study (ie. essay analysis) due to a lack of adequate representation. In order to obtain helpful qualitative representation of low scorers, another group (Low Gainers, described in detail in Chapter 6) was designed to fill the gap left in the study by the exclusion of the Low to Low group in the qualitative analysis.

Table 5.9 Low to Low Students in Subsample (n= 17, 5%) and Subsample (n=2, 4%)

	Minimum	Maximum	Mean	Std. Dev.
Sample LL N2 score pre	1.73	27.23	20.55	7.50
Sample LL N2 score post	16.28	30.93	24.50	5.14
Sample LL N2 Change	-7.87	17.39	3.9501	7.41
Subsample LL N2 score pre	18.07	20.72	19.40	1.88
Subsample LL N2 score post	32.99	34.60	33.79	1.14
Subsample LL N2 Change	12.26	16.53	14.40	3.01

5.5 FINDINGS ACROSS GROUPS

The groups originally identified in this study included four sets of categories: Gainers, Decliners, High to High scorers and Low to Low scorers. A fifth group, Low Gainers, was constructed during the qualitative analysis (see Chapter 6 for details). Table 5.10 offers descriptive statistics for each of the five groups, including the percentage of

the total sample population that each group encompasses. Knowing the range of moral reasoning development faculty might encounter in students in a program like this is helpful for responsible design and implementation of liberal arts programming. This study seeks to shed light on the likely range of moral reasoning development and typical developmental variance found in students who select into rigorous programs. Table 5.10

Table 5.10 Group statistics from total sample (n=325)

	Total Sample	Gainers <i>(N2Change ≥ 12)</i>	Decliners <i>(N2Change ≤ -6)</i>	High to High <i>Pre- and Post N2 scores ≥ 54</i>	Low to Low <i>Pre-N2 ≤ 27.24 and Post-N2 ≤ 40.75</i>	Low-Gainers <i>1stQ Pre-N2 scores and N2 Change ≥ 12</i>
N/Total	325	88	57	40	17	37
% of Total	100%	27%	17%	12%	5%	11%
Pre-test N2 mean(SD)	40.75(13.52)	30.12(12.17)	48.70(10.82)	60.27(4.51)	20.55(7.50)	22.59(5.73)
Post-test N2 mean(SD)	45.94(12.20)	50.45(10.82)	36.92(11.63)	62.44(6.28)	24.50(5.14)	47.38(7.46)
N2 Change mean (SD)	5.19(12.35)	20.33(7.87)	-11.79(6.35)	2.17(6.70)	3.95(7.41)	24.79(9.58)

* Note that the group, Low Gainers, overlaps with the Gainers group and thus is not a separate percentage of the total sample.

offers an overview of the groups identified in the study, relative to the whole sample (n=325). This information allows more adequate understanding of the likelihood that faculty might be working with students who fall into very low or very high ranges of moral reasoning development. Additionally, the data offer insight into the relative

presence of students who are apt to make great gains or whose moral reasoning development might regress. The table above reminds us that capturing a true picture of the nature of various groups includes apprehending how students' pre- and post-test N2 scores compare not only to national norms but also how they compare to each other within the total sample. In addition to the comparison of N2 pre-test, post-test and N2 Change means given in Table 5.10, a quartile analysis of the total sample provides a snapshot of within-group ranges of scores. Table 5.11 shows the total sample (n=325) via pre-test N2 scores along the x-axis and post-test N2 scores along the y-axis. In this table, we see that 42 students (13% of the sample) of the 81 students in the first quartile of

Table 5.11 Quartile Analysis Results: Total Sample subjects (N=325). Position reflects pre-test quartile (x-axis) combined with post-test quartile (y-axis) and percentage of total sample.

Post-test N2 Score Quartile		Pre-test 1 st Quartile (lowest)	Pre-test 2 nd Quartile	Pre-test 3 rd Quartile	Pre-test 4 th Quartile (highest)	
	Post-test 4 th Quartile (highest)	4 (1%)	10(3%)	22(7%)	45(14%)	n=81
	Post-test 3 rd Quartile	11(3%)	21(6.5%)	31(9.5%)	18(5.5%)	n=81
	Post-test 2 nd Quartile	24(7%)	23(7%)	21(6.5%)	14(4%)	n=82
	Post-test 1 st Quartile (lowest)	42(13%)	27(8%)	8(2.5%)	4(1%)	n=81
		n=81	n=81	n=82	n=81	n=325
Pre-test N2 Score Quartile						

pre-test scores remained in the first quartile in post-test scores. A group of 35 students (10% of the sample), moved from the first quartile to the second and third quartiles in

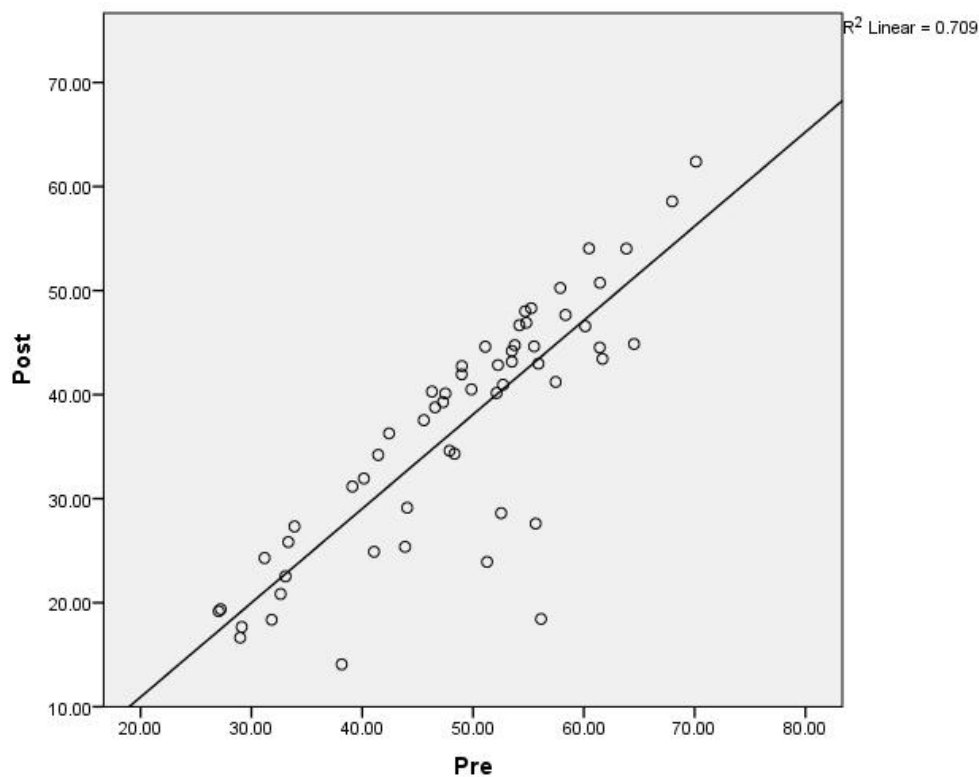
post-testing (24 students + 11 students, noted in the first vertical row) and only 4 (1%) moved all the way to the fourth quartile of post-test scores, concurring with moral reasoning literature which strongly suggests that moral reasoning is a step-wise progression. Quartile cut-points of the sample's post-test N2 scores are on average 5.41 points higher than the quartile cut-points of pre-test N2 scores, reflecting the substantial gains in mean pre- and post-test N2 scores depicted in Table 5.10. The data also highlight findings of a large group of gainers and students who retained very high scores over the course of the program. The table offers an overview of the kind of development that might be found among students in a program like this as well as the relative likelihood of moral reasoning development stasis or even decline that may be found among students.

Moral reasoning decline was an interesting phenomenon to discover in the scores from this sample. As noted in Table 5.10, 57 students (17% of the sample) posted declining scores by the end of the year. However, the mean pre-test N2 score of this group is 48.71(SD=10.82), well above national and sample norms. Twenty-eight of these decliners (49% of the decliner group) were in the top quartile of the pre-test scores, with 15 more (26%) scoring in the 3rd quartile, totaling 43 decliners (75%) scoring well above expected N2 scores in DIT pre-tests. Thus, most of the decliners were already in ranges well above expected norms and despite losses, their post-test scores tended to remain higher than norms. A scatterplot of these cases (Table 5.12) illustrates that the majority of decliners remained above national norms for college students.

As stated above, gainers represented the largest of the identified groups within the total sample, with 88 (27% of the sample) subjects identified as gaining 12 or more points (1 standard deviation of the norm of the sample mean N2 pre-test scores) by post-testing.

Fifty-four percent of the gainers came from the lowest pre-test quartile, which concurs with DIT research expectations. Only 6% of the gainers from the sample, 5 subjects, pre-tested in the top quartile. This is understandable since the top quartile was so high compared to national DIT norms that it would be unlikely to see the kind of gain (12 points or more) that defined this group in the top quartiles.

5.12 Scatterplot of Subsample Decliners



Finally, the High to High group posted quite remarkable scores in this study. Forty students, or 12% of the total sample, had pre- and post-test scores above the 54 point cut point, with a mean N2 pre-test score of 60.27 and N2 post-test score of 62.44. As noted earlier, these scores match national moral reasoning norms for PhD and Moral Philosophy graduate students. To have 12% of the total sample of first year college

students exhibiting these sorts of scores is startling and should be considered by those who are involved in the design and instruction of the program at the center of the present study. Results from the qualitative component of the study helps to flesh out these various findings and are presented in the following chapters.

Chapter 6:

Analysis of Qualitative Rubric Data and Findings

6.1 INTRODUCTION

The qualitative component of the present study set out to identify elements of student writing that reveal aspects of students' moral reasoning development and its relationship to participation in a liberal education program. As described in detail in Chapter 4, this qualitative component consisted of analysis of essays written at the beginning and end of a course by a subsample of students (n=15, in 2 class sections). As with the rest of the study sample, these students completed pre- and post-course Defining Issues Tests (DIT), a Neo-Kohlbergian measure of moral reasoning used widely for more than three decades in the area of moral development research (Rest, 1979, 1999a, 1999b) (Chapter 5 offers an overview of those quantitative, DIT findings). DIT data provided the means for identifying groups of subjects via pre- and post-test scores, test change scores and score quartile analyses. Quantitative information aided the primary researcher in identifying specific groups, including subjects with very high and very low DIT N2 scores, subjects who made great gains in N2 scores, and subjects whose DIT N2 scores declined (see Chapter 5 and later in this chapter for details on groups). Using DIT groupings, subjects were selected from the two-class subsample for essay analysis in order to further explore aspects of moral development and its relationship to course participation revealed in student writing. The following chapter gives an account of the rubric analyses involved in this qualitative component. Chapter 7 will present findings from the open, axial and selective coding of student pre- and post-course essays.

6.2 ESSAY SUBSAMPLE

From the population of 46 students who completed essays at the start and end of the class, the essays of 15 subjects (33% of the two class subsample) were selected for analysis (as described in Chapter 4). The 15 subjects included 4 Gainers, 3 Low-Gainers, 4 Decliners, and 4 High-to-High Scorers, each of whom was assigned a pseudonym (depicted in Table 6.1). To ensure a clear articulation of qualitative findings, pseudonyms were assigned beginning with a first initial that matched the first letter of the affiliated group. For example, pseudonyms used within the Decliners' group are Deb, Diane, David and Denise, while the High to High pseudonyms are Hannah, Henry, Hugh and Harold. Paper copies of all essays were scanned via OCR (optical character recognition) software to provide electronic versions of the essays for analysis. After the selection of essays was completed and essays were matched with DIT and class information, the final group of subjects whose essays were analyzed was found to include 10 female students and

Table 6.1 15 Essay Subjects chosen from Subsample of Two Classes with Pre- and Post-Course Essays (Two-class Subsample n=49, 20 males, 29 females); Essay Subject information (n = 15)

Category and Subject Pseudonyms	Parameters	N	Females	Males	Class 1	Class 2
Gainers <i>Geraldine, Grace, Gretta, Gwen</i>	N2 change ≥ 12pts	4	4	0	2	2
Low Gainers <i>Laura, Lucy, Luke</i>	T1test=1stQ; N2change ≥ 12	3	2	1	3	0
Decliners <i>Deb, Diane, David, Denise</i>	N2 change ≤ 6pts	4	3	1	2	2
High to High <i>Hannah, Harold, Henry, Hugh</i>	Pre- and post N2 score ≥ 54	4	1	3	0	4
Selected Essay Subjects Total		15	10	5	7	8

Note: See full descriptions of group parameters in Chapter 5.

5 male students, with 7 students enrolled in one of the subsample classes and 8 enrolled in the second class. A quartile analysis of the original, total study sample (n=325) is presented in Table 6.2, showing the range of pre- and post-N2 scores by quartile, with pre-test N2 scores indicated on the x-axis and post-test N2 scores on the y-axis. All fifteen essay subjects are identified (by pseudonym) within the table, placed along the axes via their pre- and post-test N2 scores. Thus, Table 6.2 reveals a student's pre-test position along the horizontal quadrants and her post-test position in the quadrants along the vertical axis. This table illustrates the pre- and post- test quartile change positions of the 15 essay writers relative to the

Table 6.2 Full Sample Quartile Analysis (n=325): Pre-and Post-Test Combined Quartiles, including percentages of total sample; 15 Essay Subjects listed (in bold) from two-class Subsample (n=15, 5 males, 10 females) identified in pre- and post-test combined quartile position

	Pre-test 1 st Quartile (lowest)	Pre-test 2 nd Quartile	Pre-test 3 rd Quartile	Pre-test 4 th Quartile (highest)	N
Post-test 4 th Quartile (highest)	4 (1%)	10(3%) Grace	22(7%) Gretta	45(14%) Hannah, Henry, Harold, Hugh	n=81
Post-test 3 rd Quartile	11(3%) Luke	21(6.5%) Lucy, Gwen, Geraldine	31(9.5%)	18(5.5%)	n=81
Post-test 2 nd Quartile	24(7%) Laura	23(7%)	21(6.5%) Deb, Diane	14(4%) Denise	n=82
Post-test 1 st Quartile (lowest)	42(13%)	27(8%)	8(2.5%) David	4(1%)	n=81
N	n=81	n=81	n=82	n=81	n=325

quartile change positions of the total sample. For instance, as noted by his position on the table, the Decliner subject, “David,” came in with a pre-test score in the third quartile but his post-test score exhibited a decline that left him in the lowest post-test quadrant. Even though one High to High scorer’s N2 post-test score declined, he remained in the highest quartile of subsample post-tests along with the three other High to High scorers. Each of the Low Gainers, Lucy, Laura and Luke, began in the first or second pre-test quartiles and ended in the second or third post-test quartiles.

6.3 ESSAY ANALYSIS

The study’s qualitative analysis design called for several cycles of essay coding. Two rubrics were used to identify within student writing, 1) evidence of moral reasoning development, and 2) aspects of the class perceived by students to be associated with moral or ethical development. The rubrics (described in detail in Chapter 4) consisted of, first, a set of moral reasoning proxy items gleaned from a thorough review of the literature on moral reasoning development research, and second, a set of course-related items generated from a review of the course’s standard syllabus and course description. The rubrics were piloted (as described in detail in Chapter 4) with a second rater to ensure rubric reliability.

6.4 ESSAY ANALYSIS: RUBRIC ANALYSES

Essay analysis began with an examination of student writing via two established rubrics. “HyperResearch” qualitative data analysis software was used to “tag” pieces of student writing that demonstrated student engagement with particular rubric items. Then, a rating system was used to rate student writing along moral development items and course-related items using a 0-3 scale (0 = rubric item absent, 1 = rubric item nominally

present; 2 = rubric item emerging; 3 = rubric item evident/active). Ratings were analyzed and compared within and across groups. Once the rubric analysis was complete, essays were reread within identified groups and coded by hand to identify patterns and themes beyond the scope of the rubrics. Patterns of phrases and word frequencies were noted in this round of open coding, general themes emerging within groups were generated in the last round of axial/selective coding. This final, selective coding provided an overview of the differences and similarities across the four groups. These cycles of rubric analysis and coding offered a unique perspective on what the range of student moral reasoning in this type of course might look like. The results of the rubric analyses and coding (untethered from rubrics) are outlined in the following sections (see Chapter 4 for full details of the rubric rating and coding procedures, including piloting of procedures with second rater). Rubric analysis for all four groups is presented first, followed by open coding themes for each group.

6.4.A Rubric Analysis by Group

6.4.A.1 Gainers' Rubric Scores

The Gainer group consisted of four subjects whose DIT N2 change scores were greater than or equal to 12 points, the equivalent of one standard deviation increase from pre-test to post-test N2 scores. Along most rubric items, subjects in the Gainer group demonstrated increased ratings in post-course essay analyses. This was expected in the case of the moral development rubric since rubric items were designed to illuminate various aspects of moral development that have been shown to be impactful in moral development literature. In the case of course-related rubric items, the data of interest

include post-essay ratings which express students' own perception of the impact of aspects of the course on their developing ideas of living the best way.

Of the 15 moral rubric items measured in the four Gainers' post-essays (for a total of 60 scores), 66% of the rated items represented increases. Only Geraldine¹ regressed along one moral rubric item. Table 6.3 offers the rated rubric items of the Gainer group, along with average ratings and increases. In terms of moral development rubric items, the lowest source of activity for this Gainer group was found along the rubric item "Diversity," operationalized here as the ability to "acknowledge and comprehend a variety of perspectives on issues and multiple worldviews" (see Chapter 4 for a complete list of operationalized definitions of all rubric items), with an average rating decline of -.4. Within the Gainer group, only one Gainer, Grace, increased .5 in her Diversity rating, while Gwen's and Gretta's ratings remained static and Geraldine's rating declined a full point. Other items that saw low ratings among Gainers were "Socratic ideal" and "context comprehension," each with an average rating increase of only .38. It is interesting to note the overlap of these three items which all deal with acknowledging the limitations of one's own understandings and context and an accompanying need to account for multiple perspectives. Moral development literature suggests that these attitudes would be associated with moral development advances, but none of these Gainers displayed a robust rating in this area. It is possible that Gainers were more preoccupied with their own developing understandings than with the limitations of those understandings. A further discussion of this will be presented below in an overview of rubric analyses integrated with themes and motifs from essay coding.

¹ Pseudonyms are used throughout the study, with initial letters of pseudonym indicative of study group, eg. a pseudonym beginning with a G is part of the Gainer group.

Table 6.3 Gainers' moral development rubric outcomes from post-course essays and change from pre-course essay (in parentheses); including average post-essay item rating and increase; **ordered from highest to lowest rating per rubric item**

Moral Rubric Item	Grace Post-essay	Gwen Post-essay	Gretta Post-essay	Geraldine Post-essay	Avg. Item Rate	Avg. Change
Principled thinking	3(+1)	3(+1)	3(+1)	2(+1)	2.75	+1
Cognitive complexity	3(+1.5)	3(+.5)	3(\pm 0)	2(+1)	2.75	+.75
Prosocial behaviors	3(+.5)	3(+2)	2(+1)	2.5(+1)	2.6	+1.13
Prosocial attitude/ Civic engagement attitude	2(\pm 0)	3(+2)	3(+2)	2(+.5)	2.5	+1.13
Ethical self-awareness	3(+2)	2(+1)	2.5(+.5)	2(+1)	2.4	+1.13
Personal & social responsibility	2(+.5)	3(+2)	2.5(+1.5)	1.5(+.5)	2.25	+1.13
Critical thinking	1(\pm 0)	3(+1)	3(+1)	2(+1)	2.25	+.75
Ethical evaluation	2(\pm 0)	2(\pm 0)	3(+2)	2(+1)	2.25	+.75
Social complexity	2(\pm 0)	3(+1)	3(+1)	1(\pm 0)	2.25	+.5
Socratic ideal	2(+1)	2.5(+.5)	3(\pm 0)	1.5(\pm 0)	2.25	+.38
Context comprehension	1.5(\pm 0)	3(+.5)	3(+1)	1(\pm 0)	2.1	+.38
Global thinking	1(\pm 0)	2.5(+1.5)	3(+2)	1(\pm 0)	1.9	+.88
Moral behavior	2(+1)	2.5(+1.5)	2(\pm 0)	1(\pm 0)	1.9	+.63
Religious attitude	2.5(+.5)	3(+2)	1(\pm 0)	1(\pm 0)	1.9	+.63
Diversity	1.5(+.5)	2(\pm 0)	3(\pm 0)	1(-.5)	1.9	0
Average rating	2.1	2.7	2.7	1.6		
Overall change	+8.5	+16.5	+13	+6.5		

The highest activity within the Gainers' ratings was found equally in four items from the moral rubric: "prosocial attitudes/civic engagement," "prosocial behavior," "personal and social responsibility" and "ethical self-awareness." *All* Gainers' scores moved positively along these four items with the same average rating increase of 1.13. The first three of these items clearly share a common concern with the benefit of others

or the community generally. The fourth, “ethical self-awareness,” adds a turn to the self and one’s core beliefs. Possibly, these Gainers are reflecting on the importance of examining one’s values in order to adequately address the mandate of “making a difference” for others. It is notable that a common exhortation to students at this university is to become “men and women for others,” and it is possible that this communal ethos sets the conditions for moral development. As mentioned above, while less change on average was seen in the categories of “Diversity,” “Socratic ideal” and “Context comprehension,” there was some variability within the Gainer group in these items. There was variability on remaining moral development items, with some gains and some losses on specific items. Notably, the item “Religious attitude” was one of the lowest rated and lowest changing items among three of the Gainers, but Gwen was deeply changed along this item. It was the only item in which one subject’s rating increased by 2 points, while all of the other three Gainers increased less than 1 point.

The second rubric was also used to analyze Gainer essays. As described in Chapter 4, these course-related rubric items were constructed via a review of moral development literature and course objectives to identify aspects of the course’s content and activities that might be impactful to student moral development. The course-related rubric items that saw the most positive activity included three rubric items, “References to texts,” “Theory and real world connection” and “Course content as interruption.” Interestingly, course-related item ratings indicating a subject’s engagement with the work of *specific* thinkers represented the lowest score change activity among these four Gainers. As depicted in Table 6.4, no evidence of the impact of “Kantian ethics” was

found in any of the four Gainers' essays, and student essays showed little or no evidence of rubric items, "Aristotelian ethics" and "Nietzschean/existentialist ethics." Gainer

Table 6.4 Course-related items rated in Gainer post-course essays and rating average; ordered from highest to lowest average rating per rubric item

Course-Related Rubric Item	Grace Post-essay	Gwen Post-essay	Gretta Post-essay	Geraldine Post-essay	Average rating
References to texts	1	3	3	3	2.5
Theory and real world connection	2	3	3	1	2.25
Course content as interruption	0	3	3	2	2
Ethical-Political theory	0	2	3	2	1.75
Ethical theory	0	2	3	2	1.75
Religious/Natural Law ethics	2	2	0	3	1.75
Common Good tradition	0	2	3	2	1.75
Nietzschean/Existentialist ethics	1	3	0	2	1.5
Aristotelian ethics	0	3	0	2	1.25
Kantian ethics	0	0	0	0	0
Total Rating Points	6	23	18	19	

mentions of course material and activities averaged < 1 point along both of these course-related items. Twelve instances of non-engagement ("0" rating) with course-related items were identified in Gainers' post-course essays, though half of those instances were associated with Grace's essay. Notably, two of the Gainers, Gwen and Gretta, wrote essays which provided 11 of the 13 highest ratings among course-related items.

Insights gleaned from both rubrics provide an interesting perspective on what development looks like among Gainers and what matters to them in terms of coursework. Gainers' concern with one's role in the welfare of others and the development of core

ethical understandings and values coincides with their tendency to use texts and course material to problematize prior ethical understandings in order to address real world concerns. They do not seem overly focused on how *theoretical* frameworks may contribute to those concerns. Gainers seem to assert personal agency and the development of personal ethics over larger, perhaps more abstract ethical systems. Of the other 7 course-related rubric items, Gainers logged only 5 instances of active engagement (a “3” rating). This sharply concentrated pattern demonstrates Gainers’ focus on the developmental tasks they consider relevant. Comparing these rubric results with those of other groups may shed light on Gainers’ seeming *disinclination* to engage with theory and with specific ethical theorists.

6.4.A.2 Decliners’ Rubric Scores

The Decliner group consisted of four subjects whose post-test DIT N2 scores declined more than a half a standard deviation from their pre-test scores. The average DIT N2 *change score* found in this group was -8.38, somewhat lower than the average N2 decrease found in the full study sample Decliners (av.-11.79) and the two-class subsample Decliners (av. -9.42). As with the rubric ratings in the Gainer group, and as indicated in Table 6.5, ratings of those subjects whose DIT scores put them in the Decliners group saw numerous areas of advanced engagement in moral development rubric items. In fact, rating *increases* among Decliners were found along two of the same moral development rubric items in which Gainers increased, “Prosocial behavior” and “Personal and social responsibility,” with average rating increases of 1.1 in each item. Decliners logged this same average increase in both “Cognitive Complexity,” a demonstrated capacity to exclude simplistic thinking, and “Ethical evaluation,” an item

signaling the subject's ability to offer a coherent articulation and defense of an ethical position. Decliners increased *most*, however, on "Principled thinking," operationalized

Table 6.5 Decliners' moral development rubric outcomes from post-course essays and change from pre-course essay (in parentheses); including average post-essay item rating and increase; **ordered from highest to lowest rating per rubric item**

Moral Rubric Item	Deb Post-essay	Diane Post-essay	David Post-essay	Denise Post-essay	Avg. Item Rate	Avg. Change
Principled thinking	3(+2)	3(+2)	1(±0)	3(+1.5)	2.5	+1.38
Personal & social responsibility	3(+2)	1.5(+.5)	2(+1)	3(+1)	2.4	+1.13
Ethical self- awareness	3(+1.5)	2.5(+1.5)	1(-1)	3(+2)	2.4	+1.00
Cognitive complexity	2.5(+1.5)	2.5(+1.5)	1(-.5)	2.5(+1.5)	2.1	+1.13
Ethical evaluation	2(+1)	2.5(+1.5)	1(±0)	3(+2)	2.1	+1.13
Prosocial behaviors	3(+2)	1(±0)	1.5(+.5)	2.5(+1.5)	2	+1.13
Critical thinking	2(+1)	2(+.5)	1(±0)	3(+2)	2	+.88
Context comprehension	2(+1)	3(+1.5)	1(-1)	2(+1)	2	+.63
Prosocial attitude/ Civic engagement attitude	3(+2)	1(±0)	1(±0)	2.5(+.5)	1.9	+.75
Social complexity	1.5(+.5)	2(+1)	1.5(+.5)	2(+1)	1.75	+.75
Religious attitude	3(+2)	1(±0)	1(±0)	2(+1)	1.75	+.75
Moral behavior	3(+1)	1(±0)	1.5(+.5)	1.5(+.5)	1.75	+.63
Global thinking	2(+1)	1(±0)	1(±0)	2(+.5)	1.5	+.38
Diversity	1(-.5)	1.5(+.5)	1(-1)	2(+.5)	1.4	-0.13
Socratic ideal	2(±0)	1(-1)	1(±0)	1(±0)	1.25	-0.25
Overall Change	18	9.5	-1	16.5		
Average Rate per subject	2.4	1.8	1.2	2.3		
Overall Change	+18	+9.5	-1	+16.5		

here as the ability to articulate universalizable principles and a desire to live by widely accepted principles, with an average rating increase of 1.4. Three of the four decliners moved from a lower rating to a “3” (highest rating) in pre- to post essays along this rubric item.

Decliners logged *decreased* ratings in post-essays along two items: “Diversity” and “Socratic ideal,” with average rating changes from pre-essays of -.13 and -.25 respectively. Decliners also saw relatively low change in “Global thinking,” indicative in this rubric of a subject’s attempt to explore the broader issues and problems of the global community. Only one Decliner, Deb, was rated above a “1” on this rubric item. These three moral development items showed not only the most *negative change* but also received the lowest ratings average by Decliners. Hence, the lowest average ratings of these three items were not simply a matter of the absence of these items in student essays, but represented *decreasing engagement with* these items from pre- to post-essays.

As for moral development item *ratings* (regardless of change), though in pre-essays Decliners rated “Diversity,” “Context comprehension” and “Ethical self-awareness” most highly, only “Ethical self-awareness” remained among the highest rated items in post-essays, along with “Principled thinking” and “Personal and social responsibility.” Items highly rated and positively changed in Decliner post-essays point to their *intentions* regarding ethics, personal responsibility and living according to widely held principles. However, the shift to *acting* on those ideas remained a low priority. The tendency of Decliners to highly value personal commitments to ideas and principles of ethics echoes Rest’s Maintaining Norms Schema and may account for where some of these Decliners may be “stuck” in aspects of this middle schema, unable to fully shift to

the Postconventional Schema. In Rest's view, the Maintaining Norms Schema involves addressing the need for ethical and institutional systems that can handle a new construal of "society" inclusive of "Others" previously excluded from the "in-group" of the Personal Interest Schema (Rest, 1979, 1999). The cognitive coordination of meeting larger social demands with functional societal systems of rules, codes, and laws for the sake of a widening circle of stakeholders is a daunting task. But the difference between *apprehending* ethical principles and waiving one's own interests to those principles is a daunting part of the shift to the Postconventional Schema. Though these Decliners exhibit increasing prosocial behavior, it does not yet rise to the level of their desire for principles and personal/social responsibility, and they are not yet sufficiently challenged by diverse and global issues that might direct them toward the Postconventional Schema.

Overall however, Decliners fared well in moral development item ratings, logging 17 highest ratings ("3") in post-essays. But two subjects, Deb and Denise, logged the large majority of those highest ratings. David on the other hand logged no highest rating and received 58% of lowest ratings of moral development items in post-course essay analyses. It is interesting to note that David's pre-test DIT N2 scores were 7.88 points lower than the Decliner average and his post-test DIT N2 scores were 6.75 points lower than average Decliner N2 scores. While the three female Decliners all posted pre-test DIT N2 scores that were well above national and whole Sample averages (all three scored ≥ 48.98 in pre-test scores), their declines left them lower than total Sample *pre-test* averages, yet still above national averages.

In terms of course-related rubric items, Decliners' post-essays engaged most actively along three course items: "Ethical-political theory," "Nietzschean/Existentialist

ethics” and “References to texts and course materials.” Echoing patterns from the moral development rubric analysis, Decliners often referenced the communal/political order called for in course texts, while also positing the primacy of the individual—a staple of Existentialist and Nietzschean paradigms (rubric outcomes depicted in Table 6.6). The combination of these seemingly disparate inclinations highlights a tension within Decliners’ developmental trajectory. While they see the need for systematic approaches to real world concerns, they also are drawn to the existentialist prioritizing of the individual. Decliners showed *least* engagement along five items: “Aristotelian ethics,” “Kantian ethics,” “Religious/Natural Law ethics,” “Common Good tradition” and “Theory and real-world connection.” It is significant to note that Decliners regularly mention having new ideas regarding the connection between political and communal order but they do not mention specific theories or authors such as Aristotle and Kant who strongly assert those connections. In general, Decliners’ essays did not exhibit evidence of course material as *interruptive* of their ideas, but tended instead to present their own ideas and then secondarily use course material as ratifying those conclusions. David’s essay in particular demonstrated this tendency. His essay included *no* mentions of course materials or course activities and only vaguely mentioned how he had changed in his thinking “this year.” He then goes on to develop a highly individualistic answer to the question of living a best way, using only his own pre-course essay as a point of reference. He claims that his ideas have developed quite a bit, but in actuality, his second answer is really a reiteration of his first answer, presented nine months earlier.

Table 6.6 Course-related items rated in Decliner post-course essays; rating average; ordered from highest to lowest *average rating per rubric item*

Course-Related Rubric Item	Deb Post-essay	Diane Post-essay	David Post-essay	Denise Post-essay	Average rating
Ethical-Political theory	3	3	0	3	2.25
Nietzschean/Existentialist ethics	2.5	3	0	3	2.1
References to texts	3	2	0	3	2
Course content as interruption	3	2	1	2	2
Ethical theory	3	1.5	0	3	1.9
Common Good tradition	2	1.5	0	3	1.6
Theory and real world connection	3	1	0	2	1.5
Religious/Natural Law ethics	2	2	0	2	1.5
Aristotelian ethics	0	2	0	1	.75
Kantian ethics	0	0	0	2.5	.6
Total rating points	21.5	18	1	24.5	

In general the Decliner group did not regularly reference authors and intellectual traditions that advocate a deep connection between ethical theory and real world considerations, but instead tended to refer to authors such as Hobbes, Nietzsche or Augustine whose texts are easily construed as asserting the priority of the individual and individual interests. Thus Decliners' essays seem to exhibit some qualities of Rest's Maintaining Norms Schema but also seemed held back by the type of self-referential tendencies and self-interest common in the Personal Interest Schema. Decliners seem to hover in Rest's Maintaining Norms Schema, with high regard for normative authorities but are unable as yet to pivot to the Postconventional Schema. In order to accomplish this

Decliners would have to reconcile the tension between social norms and individual interests, a developmental task that seems to have eluded most Decliners by the end of the course. The Kohlbergian Social Contract/Legalistic Orientation Stage coincides with Rest's initial step into the Postconventional Schema. In this stage, a subject allows that the shared norms of the community—its legal systems, duties and customs—come from a rational judgments of what is good and bad, as opposed to individual assertions of self-interest. This seems to be what Decliners cannot quite attain. While they acknowledge that they want to live in a way that seeks the good of others, or at least doesn't impinge on the rights of others, they are still held back by basic Personal Interest Schema issues.

6.4.A.3 High to High Scorers' Rubric Scores

Students whose pre- and post-course DIT N2 scores were significantly higher than the study sample *and* national norms were identified as the High to High Scorers' group. The group's scores on the moral development rubric items were quite impressive, with all four High to High Scorers rated as fully engaged (rating of "3") on 7 of the 15 items in post-essays. No post-essay rubric item saw less than an average rating of "2" (emerging). Overall change within the ratings of moral development items was moderate but this reflected the effects of the group's higher pre-essay ratings compared to those seen in other groups' pre-essays. One of the most interesting patterns noted in the High to High Scorers' ratings related to a cluster of moral development items that were problematic in Gainers, Low Gainers and Decliners (as depicted in Table 6.7). The items "Diversity," "Global thinking" and "Socratic ideal" were found to be either the lowest rated or least changed item, or both, in the post-essays of Gainers and Decliners. However, two of these items, "Diversity" and "Socratic ideal" were among the highest

Table 6.7 High to High Scorers' moral development rubric outcomes from post-course essays and change from pre-course essay (in parentheses); including average post-essay item rating and increase; **ordered from highest to lowest rating per rubric item**

Moral Rubric Item	Hannah Post-Essay	Harold Post-Essay	Henry Post-Essay	Hugh Post-Essay	Avg. Item Rate	Avg. Change
Prosocial attitude/ Civic engagement attitude	3(+2)	3(+1)	3(+2)	3(+1)	3	+1.5
Personal & social responsibility	3(+2)	3(+.5)	3(+2)	3(±0)	3	+1.1
Cognitive complexity	3(+1)	3(±0)	3(+1)	3(+1)	3	+.75
Critical thinking	3(+1)	3(±0)	3(+1)	3(+1)	3	+.75
Diversity	3(+1)	3(±0)	3(±0)	3(+1)	3	+.5
Socratic ideal	3(+1)	3(+1)	3(±0)	3(±0)	3	+.5
Principled thinking	3(+1)	3(±0)	3(±0)	3(+.5)	3	+.4
Global thinking	2.5(+1.5)	3(+1)	3(+2)	3(+2)	2.9	+1.6
Ethical evaluation	3(+2)	3(+1)	3(+1)	2(+1)	2.75	+1.25
Social complexity	2(+1)	3(+1)	3(±0)	2(±0)	2.5	+.5
Religious attitude	3(+2)	1(±0)	3(+2)	2(+1)	2.25	+1.25
Moral behavior	1.5(+.5)	2.5(+.5)	3(+2)	2(±0)	2.25	+.75
Ethical self-awareness	3(+1)	2(±0)	2(±0)	2(+1)	2.25	+.5
Context comprehension	3(+1)	2(-1)	2(-1)	2(-1)	2.25	-.5
Prosocial behaviors	1(±0)	3(+1.5)	3(+2)	1(±0)	2	.9
Average rate per subject	2.7	2.7	2.9	2.5		
Overall Change	+18	+6.5	+14	+8.5		

rated items in High to High Scorers' post-essays and "Global thinking" was found among the highest change scores, at an average increase of +1.6. The item, "Religious attitude," was also among the four items showing the greatest average increase for this group. This

item is notable since it did not appear as highly rated or highly changing for either Gainers or Decliners. The two remaining items that saw significant change for High to High Scorers were “Prosocial attitude/Civic engagement attitude” and “Ethical evaluation.” The former item was present in highest rating among Gainers (but not Decliners) while the latter was present also for Decliners (but not Gainers). Thus High to High Scorers shared with Gainers an affinity for prosocial attitudes and a desire to engage in communal issues, but were also able to evaluate ethical positions as was seen in Decliners’ rubric analyses.

These qualities were supplemented in High to High Scorers with qualities like “Global thinking” “Diversity” “Socratic ideal,” all three of which received highest ratings or highest rating increase among the High to High post-essay moral development rubrics. In fact, these three items most starkly differentiate this group from Gainers, Low Gainers and Decliners. These three items were *the lowest* rated and saw the *lowest rating change* in Decliners’ post-essays and were among the lowest rated and lowest change items in Gainers’ post-essays as well. These three items offer unique facets of the High to High Scorers. Additionally, High to High Scorers’ post-essays exhibited high ratings in “Critical thinking,” an item not highly rated or highly changed in any of the other groups.

With respect to the second rubric analysis measuring a subject’s engagement with course-related materials and activities, depicted in Table 6.8, the High to High Scorers exhibited marked differences from the other three groups. While matching levels of engagement in “Theory and real world connection,” “Course content as interruption” and “References to texts and course activities” with the other groups, this group of High to High Scorers additionally mentioned ethical and political theory regularly and robustly

throughout their essays. Their essays made wide-ranging use of theory and connected those theories with concrete, real world concerns. These students tended to highlight the

Table 6.8 Course-related items rated in High to High Scorers' post-course essays; rating average; **ordered from highest to lowest average rating per rubric item**

Course-Related Rubric Item	Hannah Post-essay	Harold Post-essay	Henry Post-essay	Hugh Post-essay	Average rating
Ethical-Political theory	3	3	3	3	3
Ethical theory	3	3	3	3	3
Common Good tradition	3	3	3	3	3
Theory and real world connection	3	3	3	3	3
References to texts	3	3	2	3	2.75
Course content as interruption	3	3	2	3	2.75
Religious/Natural Law ethics	3	0	2.5	3	2.1
Aristotelian ethics	3	3	0	2	2
Nietzschean/Existentialist ethics	0	0	3	0	.75
Kantian ethics	0	0	0	0	0
Total rating points	24	21	21.5	23	

need for systemic efforts to address issues facing individuals and focused on the interplay between individuals and institutions, persons and social orders. Essays in this group demonstrated the very sort of engagement this program would ideally seek: an intelligent grasp of material and an abiding sense of how this somewhat abstract material should and can apply to the real world.

6.4.A.4 Low Gainers' Rubric Scores

After an initial DIT analysis of the two class subsample failed to provide a sufficient number of students with very low pre- and post-test scores, a Low Gainer group was established. This group included students who *posted gains* over the year but whose DIT N2 Pre-test scores were found in the two lowest quartiles and whose gains kept them only in the second or third quartiles. Three students within this group completed both pre- and post-essays so their essays were selected for analysis. Across Moral Rubric Ratings, Low Gainers posted the most modest ratings of all four groups within the sample, including the Decliners. As seen in Table 6.9, rubric item ratings fell between 1 and 2, with only three items reaching an average rating of “2” from the group: “Ethical evaluation,” “Personal and social responsibility” and “Principled thinking.” Ratings were lowest for this group among the rubric items, “Critical thinking,” “Moral behavior” and “Global thinking.” This last item, “Global thinking,” was thus among the three lowest rated items for *every group except* High to High Scorers. “Principled thinking,” on the other hand, was rated among the top moral development items in all four groups and “Personal and social responsibility” was among the top rated or saw the most change in rating in all four groups. Low Gainers’ essays posted high change among the following items: “Ethical evaluation,” “Prosocial behavior,” “Cognitive complexity” and “Socratic ideal.” Lowest activity was seen among the moral rubric items: “Critical thinking,” “Social complexity” and “Prosocial attitude/Civic engagement attitude.”

Interestingly, Luke exhibited far greater change along moral rubric items within his essay than his two female counterparts, Laura and Lucy. Moral rubric item ratings increased 14.5 points overall in Luke’s post-course essay, while Laura’s essay posted

only a 2.5 point increase overall and Lucy's essay yielded a negative outcome of ratings on rubric items, falling .5 point overall.

Table 6.9 Low Gainers' moral development rubric outcomes from post-course essays and change from pre-course essay (in parentheses); including average post-essay item rating and increase; **ordered from highest to lowest rating per rubric item**

Moral Rubric Item	Laura Post-essay	Lucy Post-essay	Luke Post-essay	Avg. Item Rating	Avg. Change
Ethical evaluation	2(+1)	1(0)	3(+1)	2	+.7
Personal & social responsibility	2.5(+.5)	1.5(-.5)	2(+1)	2	+.3
Principled thinking	2(0)	2(+.5)	2(+1)	2	+.5
Religious attitude	1(-.5)	1(0)	3(+2)	1.7	+.5
Prosocial behaviors	2(+1)	1(0)	2(+1)	1.7	+.7
Cognitive complexity	1(0)	1.5(+.5)	2.5(+1.5)	1.7	+.7
Social complexity	2(0)	1(-1)	2(+1)	1.7	0
Diversity	1(-1)	2(+1)	2(+1)	1.7	+.3
Context comprehension	2(+.5)	1(0)	2(+.5)	1.7	+.3
Prosocial attitude/ Civic engagement attitude	2(0)	1(-1)	1.5(+.5)	1.5	-.16
Ethical self-awareness	1(0)	1.5(+.5)	2(+1)	1.5	+.5
Socratic ideal	1.5(+.5)	1(0)	3(+2)	1.5	+.8
Global thinking	1(0)	1(-.5)	1.5(+.5)	1.3	+.3
Moral behavior	1.5(+.5)	1(0)	1.5(+.5)	1.3	+.3
Critical thinking	1(0)	1(0)	1(0)	1	0
Average rate	1.6	1.2	2.1		
Overall Change	+2.5	-.5	+14.5		

In general, this group of Low Gainers posted low ratings and low change along rubric items. Luke's high increases among ratings of rubric items echoes what is found in a quartile analysis of his DIT scores. Within the DIT, Luke moved from a pre-test DIT score in the first quartile to a post-test DIT score in the third quartile. His average rating across moral rubric items of 2.1 is on par with average ratings of Gainers and some of the higher scoring Decliners, if not the average ratings of the High to High Scorers. Laura's lower average among ratings of moral rubric items at 1.6 with an overall increase of ratings of only 2.5 points corroborates her DIT scores, with pre-test DIT scores falling in the lowest quartile and post-test scores only moving into the second quartile. Lucy's scores are quite interesting, since she was the sole Low Gainer who began with pre-test DIT scores in the second quartile (higher than other Low Gainers) and ending with post-test scores in the third quartile. However, in essay analysis, her average moral rubric item rating was only 1.2, sharing with Decliner David the *lowest overall average rating* of moral items. She and David also were the only subjects within the group of 15 essay writers to post negative overall change along moral rubric items, with Lucy posting a -.5 point change overall in ratings and David posting a -1 point rating change overall.

Increases among the moral rubric item ratings, or lack thereof, also offer some clarifying information about the Low Gainer group. Within this group, ratings along all moral rubric items did not rise more than .8 points on average. Contrastingly, the Gainer group and the High to High Scorers posted average increases ≥ 1 point on five items and Decliners actually posted six items with average increases of 1 point or more. Low Gainers' essays on the other hand showed average increases $\leq .5$ along 8 moral rubric

items, while Gainers and Decliners saw only three items with such low change and High to High Scorers reported low change on only two items.

Low Gainer essays were also analyzed via the second rubric to measure this group's engagement with course materials. Low Gainers Laura and Luke both relied on course texts in answering the post-course essay question, though Lucy did not. All three of the Low Gainers mentioned a fairly substantial but only emerging sense of the course as playing a role in interrupting previous ideas about living well. The third highest rating of course-related rubric items among Low Gainers regarded "Nietzschean/Existentialist ethics," which Laura and Lucy both used with some degree of focus. Luke highlighted more keenly the connection of theory and real world issues, a connection only briefly mentioned by the female students in this group. Authors who advocate systematic approaches to ethics, such as Kant and Aristotle, and theoretical approaches such as common good theory or other ethical-political positions were largely missing in this set of essays and fell into the lowest ratings for this group. Low Gainers overall seemed more interested in approaching the question from an individualistic point of view, which is echoed in their higher rating of existentialist writings, which highlight the role of the individual. Laura and Luke referred instead to larger, systemic approaches in taking up the question, wondering if the systems of religion or ethical/political theory ought to be considered when developing a thoughtful way to live. Laura highlighted Aristotle's work, but only as regards wanting to live a life with friends in it, as opposed to utilizing his ethical theory as a whole. Luke's engagement with theory did not evince a full understanding of how it might actually play a role in forming an ethical life, seeing

theory as more as aiding us in comprehending how *others* view happiness and ethics. Lucy avoided these more systematic approaches to ethics altogether.

In general, Low Gainers' mentioned course-related items in patterns similar to the Gainer group with the main exception of a stronger impact of "Nietzschean/Existentialist ethics" (outcomes depicted in Table 6.10). However, Gainers averaged higher ratings of

Table 6.10 Course-related items rated in Low Gainer post-course essays and rating average; ordered from highest to lowest average rating per rubric item

Course-Related Rubric Item	Laura Post-essay	Lucy Post-essay	Luke Post-essay	Average rating
References to texts	3	1	3	2.3
Course content as interruption	2	2	2	2
Nietzschean/Existentialist ethics	2	3	0	1.7
Theory and real world connection	1	1	2.5	1.5
Common Good tradition	2.5	0	2	1.5
Ethical-Political theory	2	0	2	1.3
Religious/Natural Law ethics	0	0	3	1
Ethical theory	0	0	2	.7
Aristotelian ethics	2	0	0	.7
Kantian ethics	0	0	0	0
Total Rating Points	14.5	7	14.5	

course-related items, averaging 16.5 rating points as opposed to the Low Gainers who averaged only 12 rating points overall on course-related items. Low Gainers' reliance on Nietzschean/Existentialist ethics is unsurprising, since it appeals to the sort of individualism that might be detected in the DIT. Low Gainers' disinclination for theory

may signal personal interest schema concerns or the sort of instrumental-relativism of lower schema thinking from a neo-Kohlbergian point of view. Like Gainers, the Low Gainer group ended up preferring personal ethics formats over theoretical approaches. However, Laura and Luke both indicated emerging consideration of systematic, structural approaches to ethics even though their individualism eventually trumped these theoretical considerations.

6.4.B Rubric Patterns across Groups

A comparison of rubric ratings across groups provides a useful overview of the facets of moral reasoning development of these four different types of students as well as the kinds of course materials and activities that may be impactful for them. Discovering which “pieces” of moral reasoning are relatively active or inactive within particular student groups is illuminative of what instigates and/or stifles progress in this area of development.

Unsurprisingly, the ratings of High to High Scorers were markedly higher than the ratings within other groups. In fact, the lowest rating in that group, “Prosocial behavior” received a rating average that was *as high as the highest average rating* of the Low Gainer group, as depicted in Table 6.11. The moral reasoning development item that was rated highest for each group was “Principled thinking,” an item that signals a subject’s ability to articulate universalizable principles and expresses a desire to live by widely held principles. The subjects within all four groups clearly prioritized this endeavor. In general, we may say that the task of trying to identify solid and accepted principles to live by is a commonly shared goal among all subjects. Similarly, “Cognitive complexity,” defined in this rubric as demonstration of a subject’s capacity to exclude

Table 6.11 Moral Development rubric items ranked from highest to lowest (1-15) based on average rating per item by group, 0 (moral development item absent) -3 (moral development item highly engaged)

Rank	Low Gainers avg. Ratings	Decliners' avg. Ratings	Gainers avg. Ratings	High to High Scorers' Avg. Ratings
1	Principled thinking 2	Principled thinking 2.5	Principled thinking 2.75	Principled thinking 3
2	Personal & social responsibility 2	Personal & social responsibility 2.4	Cognitive complexity 2.75	Personal & social responsibility 3
3	Ethical evaluation 2	Ethical self-awareness 2.4	Prosocial behaviors 2.6	Prosocial attitude/ Civic engagement attitude 3
4	Cognitive complexity 1.7	Cognitive complexity 2.1	Prosocial attitude/ Civic engagement attitude 2.5	Cognitive complexity 3
5	Prosocial behaviors 1.7	Ethical evaluation 2.1	Ethical self-awareness 2.4	Diversity 3
6	Religious attitude 1.7	Prosocial behaviors 2	Personal & social responsibility 2.25	Critical thinking 3
7	Social complexity 1.7	Critical thinking 2	Critical thinking 2.25	Socratic ideal 3
8	Context comprehension 1.7	Context comprehension 2	Ethical evaluation 2.25	Global thinking 2.9
9	Diversity 1.7	Prosocial attitude/ Civic engagement attitude 1.9	Social complexity 2.25	Ethical evaluation 2.75
10	Prosocial attitude/ Civic engagement attitude 1.5	Social complexity 1.75	Socratic ideal 2.25	Social complexity 2.5
11	Ethical self-awareness 1.5	Religious attitude 1.75	Context comprehension 2.1	Religious attitude 2.25
12	Socratic ideal 1.5	Moral behavior 1.75	Global thinking 1.9	Moral behavior 2.25
13	Global thinking 1.3	Global thinking 1.5	Moral behavior 1.9	Ethical self-awareness 2.25
14	Moral behavior 1.3	Diversity 1.4	Religious attitude 1.9	Context comprehension 2.25
15	Critical thinking 1	Socratic ideal 1.25	Diversity 1.9	Prosocial behaviors 2

simplistic problem solving, was rated in the top four rubric items in all groups. The prioritizing of this item across groups echoes the relatively high DIT N2 scores found in the total study sample (see Chapter 5 for details). Note that this item is *not* simply intellectual ability, which this study did not seek to correlate since DIT research strongly suggests that moral reasoning does not reduce to intellectual ability. “Personal and social responsibility” was yet another item that found its way into the highest ratings of many of the groups, with a first highest average rating among High to High Scorers and Low Gainers, and a second highest rating among Decliners. Only Gainers rated this item, defined as the seeing oneself as having agency in the welfare of others and the community, sixth out of fifteen items. However, it is interesting to note that Gainers’ second highest rating went to “Prosocial behavior” or adverting to social action intended for the benefit of others without anticipation of personal reward. None of the other groups included this item in their top average ratings. In fact, it was the lowest rated item for High to High Scorers. It would seem that while most subjects comprehend their own role in the good of others and society generally, only Gainers made moral or ethical action a priority over ideas of moral obligation.

The area of clearest contrast regarded items “Global thinking,” and “Diversity,” which both deal with the degree to which a subject acknowledges and seeks understanding of global issues and to try to see them from different perspectives. As depicted in Table 6.12, “Global thinking” was among the three lowest average ratings for Low Gainers, Decliners, and Gainers, but was most among the top ratings by High to High Scorers. Similarly, “Diversity” was among lowest ratings for Gainers and Decliners

Table 6.12 Moral development rubric items ordered alphabetically, with group average ratings per item

Moral Development Rubric Item (in alphabetical order)	Low Gainers Avg. Rating	Decliners Avg. Rating	Gainers Avg. Rating	High to High Scorers Avg. Rating
Cognitive complexity	1.7	2.1	2.75	3
Context comprehension	1.7	2	2.1	2.25
Critical thinking	1	2	2.25	3
Diversity	1.7	1.4	1.9	3
Ethical evaluation	2	2.1	2.25	2.75
Ethical self-awareness	1.5	2.4	2.4	2.25
Global thinking	1.3	1.5	1.9	2.9
Moral behavior	1.3	1.75	1.9	2.25
Personal & social responsibility	2	2.4	2.25	3
Principled thinking	2	2.5	2.75	3
Prosocial attitude/ Civic engagement attitude	1.5	1.9	2.5	3
Prosocial behaviors	1.7	2	2.6	2
Religious attitude	1.7	1.75	1.9	2.25
Social complexity	1.7	1.75	2.25	2.5
Socratic ideal	1.5	1.25	2.25	3

but among the highest rated items for High to High Scorers. Indeed, this item was rated more than a full point higher by High to High Scorers than any other group (next highest was the Gainer group) and more than doubled the rating of Decliners on that item. “Socratic ideal,” an item defined as acknowledging the limitations of one’s own understanding, contexts and experiences, was low for Decliners and relatively low for Gainers and Low Gainers, but very highly rated by High to High Scorers. Finally, Low Gainers’ lowest average scores came in on the rubric item “Critical thinking,” defined as

a demonstrated capacity to synthesize ideas, images, skills to address problems in innovative ways, while this item emerged as one of the highest rated items by High to High Scorers.

Analysis via the course-related item rubric produced findings across groups that offer several fruitful areas for consideration. As shown in Table 6.13, High to High Scorers were rated on average the highest possible rating (3) on four course-related items: “Ethical-political theory,” “Ethical theory,” “Common Good tradition” and “Theory and real world connection.” That is, all four students in the High to High Scoring group mentioned or made use of theoretical approaches to ethical, political and common good issues and demonstrated a relatively advanced or nuanced grasp of those approaches in their post-course essays. They also regularly focused on the connection of those theoretical considerations with concrete and contemporary real world issues. In contrast, most combinations of these items were lower in the course related item ratings among the three other groups, with the sole exception of Gainers, whose essays demonstrated engagement along the rubric item “Theory and real world connection” as well.

All four groups saw relatively high ratings regarding their references to texts and acknowledging the role of the course as interruptive of previously held ideas and beliefs. Three groups shared a common lowest rating (0) on the author/text-specific item, “Kantian ethics,” with the Decliners displaying some, albeit low, activity on that item. This low activity may reflect students’ rejection (or some Decliners’ acceptance) of a perceived rigidity within the Kantian ethical system. From a neo-Kohlbergian point of view, this rejection of the seemingly inflexible set of Kantian ethical absolutes makes sense as subjects attempt to reconcile complications encountered in the Conventional

stages. Kant's call for grounding one's morals in highly rigid formulas resonates with a Kohlbergian law and order-based ethics (Stage 4). Rest's Maintaining Norms schema would also encompass this moral mode. Aristotle's virtues-based ethics rates slightly higher engagement in students' post-course essays, though not by much. Students

Table 6.13 Course-related rubric items ranked from highest to lowest (1-10) based on average rating per item by group

Rank	Low Gainers average ratings	Decliners average ratings	Gainers average ratings	High to High Scorers average ratings
1	References to texts 2.3	Ethical-Political theory 2.25	References to texts 2.5	Ethical-Political theory 3
2	Course content as interruption 2	Nietzschean/Existentialist ethics 2.1	Theory and real world connection 2.25	Ethical theory 3
3	Nietzschean/Existentialist ethics 1.7	References to texts 2	Course content as interruption 2	Common Good tradition 3
4	Theory and real world connection 1.5	Course content as interruption 2	Ethical-Political theory 1.75	Theory and real world connection 3
5	Common Good tradition 1.5	Ethical theory 1.9	Ethical theory 1.75	References to texts 2.75
6	Ethical-Political theory 1.3	Common Good tradition 1.6	Religious/Nat. Law ethics 1.75	Course content as interruption 2.75
7	Religious/Nat Law ethics 1	Theory and real world connection 1.5	Common Good tradition 1.75	Religious/Nat. Law ethics 2.1
8	Ethical theory .7	Religious/Nat . Law ethics 1.5	Nietzschean/Existentialist ethics 1.5	Aristotelian ethics 2
9	Aristotelian ethics .7	Aristotelian ethics .75	Aristotelian ethics 1.25	Nietzschean/Existentialist ethics .75
10	Kantian ethics 0	Kantian ethics .6	Kantian ethics 0	Kantian ethics 0

mentioned his ideas on relationship and friendships more regularly than the central ideas of his ethical treatise—the virtues needed to live a happy life. Similarly, authors and texts from the religious and Natural Law traditions received relatively slight mention.

The most fascinating course-related rubric comparison across groups emerged in different prioritizing of “Ethical theory” or “Nietzschean/Existentialist ethics.” As stated above, all students mentioned references to texts and the course as interruptive of their own ideas. However, the actual course material and ideas from texts that different groups highlighted presented a startlingly different picture across groups. While Low Gainers and Decliners both highlighted “Nietzschean/Existentialist ethics” fairly regularly, rating on this item among Gainers and High to High Scorers was relatively low. These latter groups showed greater engagement with and mentioned much more regularly course material regarding ethical theory and ethical-political theory. Gainers’ and High to High Scorers’ partiality for comprehensive ethical theory *combined with* their prioritizing of the connection of theory and real world issues would be rewarded in a neo-Kohlbergian context which prizes this sort of higher order thinking and movement to connect ethical thinking to ethical action. Their preference for theoretical and systemic treatments of ethical issues over the individualist framework of Nietzschean/Existentialist ethics reiterates their commitment to moving beyond the scope of individual interests and “in-group” preferences. It also echoes High to High Scorers’ and Gainers’ predilections for large-scale, systemic thinking in attempting to resolve the limits of simplistic law-and-order arrangements.

Chapter 7:

Essay Coding Analyses and Findings

7.1 ESSAY ANALYSIS: THEMES/MOTIFS DETECTED IN OPEN AND AXIAL CODING

A first cycle of open coding was completed on all pre-course and post-course essays from the 15 subjects chosen from the subsample. Words, phrases and sentences were highlighted and saved via Hyperresearch software and tagged to associate each item with moral development or course-related rubric items. Essays were subdivided into groups (details of groups can be found in Chapter 4) according to the study's parameters and open codes were then merged to produce a set of axial themes for pre-essays and post-essays for each group. Thus, two sets of axial codes were identified for each group, including pre-essay themes and post-essay themes (eg. Gainers' pre-essay themes and Gainers' post-essay themes). A final round of selective coding was completed using these axial themes in order to glean the core or central themes of each group for comparison across groups. The following is a report of the findings from these analyses.

7.1.A Themes within Groups

7.1.A.1 Gainers' Themes

Several common themes emerged in open and axial coding of the Gainer essays. In pre-essays, all four Gainers emphasize the challenge of trying to answer a question like "What Is the Best Way to Live?" during the first week of college. Gretta² reflects that "now that I have entered into a new chapter of my life...I have found myself in more diverse social situations in the first week of college than in my whole high school

² Once again, pseudonyms are used throughout the study, with initial letters of pseudonym indicative of study group, eg. a pseudonym beginning with a G is part of the Gainer group.

career....I find myself missing the comforts of home and the security of the people I know so well.” Gwen notes that “sometimes happiness is found, and sometimes the search continues...what we perceive happiness to be is constantly changing,” echoing Grace’s stated desire for “a foundation that is constant, when everything else around me is constantly changing.” The theme of searching for a purpose or foundation in the midst of a changing horizon is pervasive in the Gainers’ pre-essays and is quite different from the sort of security issues found in the essays of the Decliners. While Decliners focused on the need for physical and familial security, Gainers focus instead on a desire for a personal foundation and sense of purpose and meaning. Themes of anxiety about uncertain and shifting personal foundations were largely missing in the essays of other groups, most clearly absent in the High to High pre- and post-essays.

Gainers regularly reported in pre-essays that finding an answer to the question of a good and happy life would entail hard questions and arduous searching, with Gretta noting that “it takes bravery to live a reflective life...[taking] a lot of perseverance to keep reassessing your life and keep trying again to find the place where you belong.” These four pre-essays were full of words, phrases and ideas about challenges, striving, pursuing meaning, gaining foundations and finding purpose. Unlike any other group in the study, most Gainers posed numerous questions within their own essays. Geraldine suggested that the process of finding purpose would be “lengthy and open-ended” and would raise hard questions such as, “why am I here, and what am I doing?” Gretta highlighted the need for good and hard questions that would push her to reflect more deeply, at one point in her essay listing five questions in a row including, “Are these people like me? Do they share my values?.... These questions drive me insane...but now

that I am putting these questions to the test, I find that they help me on my road to finding my place....” She goes on to note that “[i]t is also a challenge to ask the right questions,” and not get caught up in the wrong, shallow questions that don’t lead one further toward meaning and a sense of purpose. Grace asserts confidently that the best way to live is “searching for [a] purpose, finding it, and then pursuing it.” This Gainer preference or predisposition for questions and challenges is echoed also in the Gainers’ expressed need for openness in this new college experience. Geraldine notes, “I want to push myself to try new things and meet new people,” while Gretta adds, “these answers don’t come easily...but the fact is at least I keep asking them.” Finally, Gwen sums it up this way: “I do not want to simply follow the crowd....I will aim to make my own decisions about how to make my life the best possible for me.”

It is not surprising that Gainers’ post-essays reiterate the struggles inherent in the search for *sustainable* answers to life’s challenging questions, for purpose and meaning. Three of the Gainers highlight how hard this process is: “Many things have changed [for me] in this time period...but one thing has not changed, I still understand the importance of asking meaningful questions” (Gretta); “It’s hard to determine when exactly this question is answered...I do not think the good life is free of moments when you doubt this purpose or it seems lost...but you wrestle with asking yourself tough questions...” (Grace); “If happiness were easily attained, it would just as easily be lost” (Gwen). All four Gainers describe the *search* for authentic meaning, purpose or happiness to be central to finding a best way to live and describe that search as a *process*, something to be continuously sought, that is only just begun. Two of the Gainers stress in particular the challenge of this process, citing the cost of the search for knowing who you are and what

you truly believe and stand for. Gwen notes that suffering may aid us in gaining perspective on what really matters and while Gretta notes a universal ability to gain a life of goodness, she asserts that we “just need to be brave and truthful enough to do it.” These young women offer Viktor Frankl and Martin Luther King, Jr. as examples of figures who chose not to avoid the struggle that *is* a deep and sustained examination of the self and the state.

7.1.A.2. Decliners’ Themes

While Gainers’ pre-essays focused on the precarious and uncertain nature of being a first year college student with many questions and much to learn, students in the Decliner group reported things in a quite different light. Their pre-essays were not peppered with questions, as the Gainers’ essays largely were. Instead, their essays took the the form of sure and solid opinions with not a single question listed in all four essays. These unquestioning voices revealed a dualism that was missing from the other groups. Essays and paragraphs began not with questions or statements qualified by “I think that...” or “It may be that...” as found in essays of other groups, but with strong unqualified statements like “A person must...” or “The primary focus in any person’s life must be....” These four Decliners exhibited a distinct air of confidence in their capacity to answer the question about living the best way, though almost all asserted the need for security and balance and several mentioned anxiety about making mistakes, wasting time, or letting restrictions hold them back. In general, these essays were profoundly different than their Gainer counterparts in both tone and content.

All four Decliner pre-essays emphasized the need for balance. This was consistently highlighted by the four Decliner subjects. Diane focused much of her

reflection on a pyramid of needs (without citing Maslow, though clearly using his theoretical framework), spending more than one quarter of her paper discussing the human need for food and shelter and the effects of these needs on stress levels. She further argues that a fulfilling life would entail the “delicate balance of acting on one’s whims while maintaining a clear conscience” once these basic needs are met. Deb is similarly concerned with balance, likening living the best way to a high-stakes poker game. Her professional-gambler Grandfather taught his family that life is about knowing “when to hold a good-looking hand” and she chastises herself for youthful attempts to reach too far, too soon, seeing these as “simply stumbles in my pursuance of a balance between safe and risk-taking behavior.” She concludes with the reflection that we should recall “which gambles and refrains have worked and failed in the past, so that [we] can find the balance between the two and evolve into better human beings.” Subject David concurs, positing in his pre-essay that problems from his youth resulted from not understanding how to find the balance between perfectionism and being too carefree, finding now that “having a healthy mixture of both has...allowed me to learn exactly how I feel academics, and most of life in general, should be approached.” His conclusion that “being happy means finding that perfect balance between striving to do great things and having fun every day” aligns with the fourth Decliner, Denise, who lists five factors that “when combined, all...create a great balance between work and play” which is central in her picture of the good life. The theme of balance among Decliner essays was one of the most consistently found patterns among all of the essays. This concern may be connected to subjects’ need for a sure and secure foundation in the midst of the important transitional moment of starting college.

Decliners are keenly aware of factors that may threaten this delicate balance. For the most part, they view insecurity and external influences as the primary disruptors of perfect balance. Diane suggests that “life is riddled with numerous restrictions...two of the most powerful and influential...are gravity and conformity” which both cause physical and mental stagnation and atrophy. Diane’s conflation of conformity with the physical reality of gravity intimates a certain sense of powerlessness in the face of social convention or conformity. “Comfort is a huge factor,” claims Denise, who recognizes that finding the right balance between ambition and success will determine her capacity to have and do what she wants. She maintains that “with success, I will be able to afford a nice sports car and a high-end home...and teach in a third world country like Rwanda,” a goal which will satisfy both her passion to help others and her “interest in adventure.” Deb’s framework of life as a high stakes card game caused her to “[fear] that all my hard work would go to waste” at certain times in high school when her success seemed unsure. Decliners seem to perceive many external threats to the balance they crave and the goal of neutralizing those threats is apparent throughout their essays.

One of the most unique and consistent characteristics of the Decliner group pre-essays was what might be called a backward-looking quality. In three of the four essays, family and high school events and friends figure prominently, while the fourth essay focuses on the Maslow’s hierarchy of needs (presumably learned in a high school class). While students in the other groups mention high school and high school friends in passing, the Decliners were the only students to mention specific family members within their essays. Deb, David and Denise all highlight family and high school friends/events as central in their pre-essays, with Denise claiming that “[f]amily is a huge factor in my

quest for the perfect life....They are, in a sense, my foundation, so if I ever question myself, I can always go to my parents. They would give me advice on certain matters and maybe even bake me cookies and yummy desserts when I feel down!" Having "friends who know me and care about me" from high school is central to David's sense of moving forward to find out who he truly is. This quality of looking backward to family and high school makes sense in light of the need for balance and security expressed by this group of students. The Decliners' self-assurance is clearly premised on solid family ties but these students seem preoccupied with looking back to that foundation rather than looking ahead or to their new settings for help in defining what a best way to live might consist in.

By the end of the year-long course, Decliners' post-essays presented with very high levels of abstraction in characterizing what was needed for a best life. This was especially true in post-essays of the three female Decliners. Diane wrote eloquently and yet *very* abstractly on the need for authenticity in a successful life, noting within the essay the absurdity of trying to achieve it and the impossibility of ever fully understanding ourselves or others. She suggests that there is an "intellectual tension necessary to sustain the authentic self [and]...to maintain the emotional conditions needed to allow the individual to move fluidly, purposefully and meaningfully in life." She goes on at length in a somewhat robotic tone about the temptations of the mind that persuade us that we can find meaning, such that we "go through the motion of living, but...without purpose and under false pretense...nonsensical and void of meaning." Her idea of authenticity seems starkly alone and purely intellectual, though she admits that an authentic self "will be naturally inclined to live with others in a community and become a member of a

political system.” However, this persistently high abstraction belies Diane’s call for authentic community. The form of community she describes seems rather like drudgery, sounding neither appealing nor workable, and Diane herself concludes that society seems to inevitably create a psychological paranoia “in which one may never feel truly secure in her position in life.” Denise, too, concludes in her post-essay that the best life must be lived in a “society which comes at the cost of self-contentment and freedom” in which citizens are *required* to be active thinkers and would be compelled to treat those in lower classes (seen by her as an unfortunate but necessary aspect of society) with dignity and respect. She sees this working more effectively under the auspices of a highly conceptual belief in a God which is a “perfect being you can strive to be like and look up to.” Denise’s post-essay offers abstract ideas about establishing religious, political and social orders “full of enlightened individuals” who surrender their personal freedoms for the sake of communal order. But she seems half-hearted in this assertion. In a postscript to the essay, she adds, “I mean honestly, my dream world would have unicorns and rainbows in it. But since that is not practical, I guess I can settle for this one!” In the case of Deb, who drops her previously-used gambling analogy in her post-course essay, high abstraction also moves along religious lines. She notes that when she “hit rock bottom” during the year, she “prayed like I never had before...and I found self-love through God...now my life has meaning because of God’s love.” She concludes with a rather idealized view of how this would work, maintaining that, “life has meaning because of God’s love....[e]ach person needs meaning, and the best and only way to find that is through the meaning that God’s love instils in each human being.”

Deb's shift from viewing life as a high-stakes card game to a seeming surrender to God was certainly connected a traumatic event during her first year of college which she reported in her essay. It is notable that two of the four Decliner essays included references to a significant emotional loss. David, whose decreased post-test N2 scores were the lowest in the subsample of essay writers, seems to "double down" on egoistic themes in his post-essay in which he refers to a difficult ending of an important relationship endured during the year. Rather than moving toward the high abstraction of the other Decliners, David writes about the experience of unrequited love which made him question himself. He considers his first essay to be very philosophical (though it wasn't particularly so), reiterating and recommitting to his decisions in high school to find balance between perfectionism and being relaxed. For several pages, he recounts his reaction to living through the break up, determined now to find what is missing in his life. He describes losing weight and rebuilding his "muscle mass" in ways that "he hadn't imagined that he could," positing that in this physical transformation "my shallow motivations evolved into feelings of accomplishment and confidence." He writes extensively about capturing lasting happiness, noting that "we are frustrated and unhappy when a person does not care about us the way we care for him or her, or similarly when our devoted time into an action or skill does not pay off the way we had desired." Unsurprisingly, his essay is entitled "Reciprocated Love" and his conclusions tend toward a romanticized view of love that is willing to sacrifice everything for another person. Deb's essay focuses on her attempts to find meaning in her grief following a traumatic event. She looks back on her initial, self-destructive reactions to this trauma concluding that the only way forward is to move beyond egoism to find a supportive and

loving community. She writes, “[e]ven though my four year old self tells me that it is, everything is not about me. Me, ME! But the I is important because when we start to correct ourselves and self-destruct less we can begin to offer the world our greatness.” Deb recounts great satisfaction in having been able to help a young girl while serving as a summer Bible camp counselor. She points out that the camp staff and everyone in her life who has loved and cared for her had done so out of a deep and abiding religious faith. In a time of trauma and grief, Deb’s post-essay is clearly marked by an intense search for a firm foundation to keep her from falling into a profound existential and moral abyss. She moves toward a communal model that first offers love and comfort, from which moral and ethical goodness will follow, “it is evident that a community’s job is to show others the right way. If we teach love, and preach love, people will be better.”

Decliners’ post-essays do not entirely lack the sort of confidence and assuredness of their pre-essays. Though these essays do not communicate the high self-confidence of their pre-essays, Decliner post-essays still do not pose questions or refer to the issue of differing opinions on the central question of the best way to live, as many other essay writers did. Their use of textual references from the course consists almost entirely of noting how various thinkers support their own conclusions. The general tendency not to look to other ideas and to diverse opinions about the best way to live, accompanied by their wariness of external influences (demonstrated clearly in pre-essays) illuminates the sort of developmental retreat that many young adults experience when faced with personal and intellectual challenges. Though many of these Decliners acknowledge to the needs of those around them and of communities generally, their attempts to answer those needs come in highly abstract formulas. In thinking about how these students are or are

not moving forward in their preference for Postconventional Schema thinking, we might consider what sorts of basic and sometimes traumatic Personal Interest Schema concerns are drawing them back to lower schemes of moral development.

7.1.A.3. High to High Scorers' Themes

The essays written by the four students in the High to High Scorers subsample were remarkable in a number of ways. High Scorers' pre-course essays were articulate and presented a variety of themes. Rather than dwelling on the daunting challenge of the essay question as the Gainers had, or confidently approaching the question with certitude as the Decliners tended to do, the High Scorers took the approach of the pure or true relativist. These four high scoring subjects posited in their September essays that while demands of the world were important to consider, the question of the "Best Way to Live" could only be answered *by the self, for the self*. Henry begins, "When I think of the good life, no formula springs to mind....Each life is by its nature unique, indeed...facets of human existence render it not conducive to standardization or formulization....Any answer, then...will be inherently inadequate....To me then, the good life must be defined in the abstract." Harold concurs with this principle, writing that "[W]hen answering the question 'What Is the Best Way to Live?' it is important to remember that the question refers to creating the best life for only that person, not anyone else. While it is true that helping other people is an important part of life, it does not benefit the subject of the question at all."

This focus on the self, however, does not take the form of self-absorption or self-centeredness in High Scorers' pre-essays. Rather, their individualism seems to have an existentialist quality, focused on themes such as viewing life as a whole, living one's life

in a fully engaged way and authentically participating in life and in the world. Hugh asserts that we should always be open to change and growth, reflecting primarily on the aphorism, “[When] we cease to grow, we cease to live.” He concludes that openness and a willingness to be corrected and redirected by others are essential to living the best way since, “change is the very essence of what it means to be human.” Hannah echoes this chord, asserting that when “we become open to self-discovery, we become free to find our own happiness and we lose the clutter in our lives to find clarity.” For her, this getting to the essence of things beyond “the clutter” can “direct your life’s journey toward the things you find beautiful so that those moments when your life seems clear and in order...can turn into a lifetime of joy.” She mentions the sunsets of Utah as the thing she misses most from home, not because they remind her of home or family, but because they have the capacity to plunge her into a more deeply felt experience of the present moment. Henry’s version of this sentiment claims that “what matters is one’s disposition towards their own life....Living the good life is thus a decision...a decision to seek one’s own truth, one’s own priorities, one’s own values...unique to the soul who has crafted [them].”

These rather cerebral construals for the most part remain fairly abstract in the group’s pre-essays, though High Scorers often allude to the need to ground their highly existentialist reflections in the real world. For instance, Henry notes that if one doesn’t figure out how to truly live his own life to the fullest, all the good he might want to do for others still would not be the best way to live, since “[at] the end of your life, you could have made a multitude of positive changes to the world, but if you did not enjoy a single second of it...what did you accomplish in life?” He concludes that finding love and

taking advantage of every possible opportunity offers the most benefit to the self and others. Hugh is more explicit in his desire to make a difference in the world, claiming that “what is most important to me is effecting change in the lives of those around me and in the community as a whole, in the spots that I believe need that change.” Despite these intimations of engagement with others and the world, however, the High Scorers’ pre-course reflections remain relatively conceptual and nonconcrete. With one brief exception in Hannah’s pre-essay, none of the High Scorers mention home, family, high school or any perceived challenges of their present transition to college. These four subjects seem oriented to a larger, more abstract future and consider this essay question to be about life as a whole, seemingly without immediate implications. For the most part, they conclude that the question of living a best life is largely a progression, a *process* of decision-making or a journey that every individual must take up.

By their post-essays, the High Scorers reassert this holistic approach but now more explicitly link its demands to society as well as to the self. Hannah, who in her pre-essay posited enjoying sunsets as an example of how to live more fully in the moment, now sees her original insights as “incomplete and vague” but still pointing to something essential about those moments when she felt or perceived a *rightness of being*. She writes of being “most happy when my life felt *in order*.” But she adds now that “order and harmony are also the proper *telos* [aim] of society and should be treated as a societal end” since “society functions in the best way possible when it is rightly ordered, because whatever society values as the greatest good will affect all other things.” For her, understanding the “interconnectedness of people and the reality that each action of the individual affects the whole of society” provides for us a sense of the “holistic reality of

society.” Harold also writes in very similar terms describing a “holistic individual” to be “truly good, for [he] strives to eliminate injustice and is genuine in the sense that it does not turn away from legitimate questions concerning the welfare of society.” He expresses a desire to strive for virtue by “addressing relevant issues facing me and the community around me” even though “this course of action has no benefit on [his own] social status.” All four of the High Scorers’ post-essays focus explicitly on the close link of individual and societal flourishing and emphasize the need for holistic social orders and systems. This group, much more than the other groups, takes a global perspective on the question of the best way to live and refer regularly to the interconnection between social and personal values. Henry posits that “while it is the individual’s job to create her own meaning, it is society’s job to create space for meaning to be pursued....We all have inherent dignity, which must be recognized both by our society and ourselves if we are to live the best way possible.” Hugh suggests that “human harmony” demands that we fight against capitalism’s “dehumanizing effects” that impact “aspects of our lives that previously resided outside the economic realm.” His concern that individuals may not be able to truly flourish within disorienting or corrupt social orders resonates with the global concerns of all four High Scorers.

Indeed, while High Scorers’ post-essays were replete with issues of global scale, they also voiced concrete and particular concerns. While Hugh’s post-essay includes many textual references, he focuses his essay on a documentary about the economic collapse of 2008 which the class watched as an assignment. Like all of the High Scorers, Hugh admits that there are no easy answers to big problems facing society, noting that in the case of the economic meltdown, “good people making supposedly good decisions

resulted in the unjust economic ruin of so many.” He goes on to express his ambivalence about the “moral neutrality” of systems and institutions when so many people have to bear the brunt of systemic breakdowns. Henry more stridently declares that “systems reduce the individual” and arbitrarily define values and meanings. He asserts that this is why he works on political campaigns and argues that “we must be honest with ourselves about the root of the systems that underlie our society and recognize where we degrade rather than bolster the dignity of the individual.” Closer to home, Harold notes the regular disparity in college life between a “lifestyle with the appearance of goodness and one that is truly good” and states that “I am beginning to realize that the life I ought to live might not look much like this one.” Two full pages later, he revisits this theme in his conclusion, writing that, “it seems unlikely that the guy slapping girls’ butts at a party will go on to have any sort of ability to solve issues of injustice in the world.” This constant interchange between the flourishing of individuals and social systems is a clear mark of the High Scorers’ post-essays, indicating the propensity of the students in this group to pivot seamlessly from questions of individual goods to common goods. This clearly is the sort of ability sought in the DIT, which prizes one’s capacity to hold in adequate tension abstract values of justice on the one hand and the particular and immediate needs of individuals on the other.

Rest’s construal of the Postconventional Schema ratifies High Scorers’ conclusions that while global and systemic problems may seem daunting, we should not fall into resignation or despair. These four students were surprisingly optimistic about what individuals and societies might accomplish. Hugh hopes that we might be called “to build a new economic paradigm, a system in which students read for the sake of gaining

knowledge and in which being a good banker means looking after the money of others rather than gambling with it.” As Henry adds, “my conception of the excellent life is predicated upon the simple notion that humans have the capacity to live excellently...to take seriously the truth of our lives without succumbing to despair and meaninglessness.”

Finally, it is helpful to note that all of the four High to High Scorers highlight the importance of systems that work for the sake of the dignity and welfare of individuals. All four of these students stress the need for social order that encourages personal flourishing and social harmony. Their calls to create political and economic systems that promote goodness, respect and dignity were tinged with religious overtones, but for the most part, these students were restrained in their appeal to religious foundations. Hannah, who purports to have no particular religious commitment, finds value in the kind of “agapic love” and cosmic purpose that she has heard about from faculty and administrators. She balks at claiming any specific belief or insight into God, but sees religious ideas as helpful in recognizing “that our actions matter beyond just ourselves and affect the absolute and infinite scheme of things...and thus each individual’s actions have great significance in the big picture.” Still, these High to High Scorers seem to move from a version of pure relativism in pre-essays to exactly the sort of moral complexity that Rest and Kohlberg were seeking to identify. These students display an ability to move seamlessly between the demands of the individual and society, expressing nuanced understandings of the ways that social, economic, political and religious contexts intersect with concrete needs of individuals. For these students, questions of justice and morality are framed with an eye to how those may be worked out in people’s lived experiences and they thus assert that social and political orders ought to elevate

human interconnection and foster greater respect for human living. These are heady assertions from 18 and 19-year olds. Their insights into the reciprocal relationship between macro- and micro-moral demands coincide quite precisely with the characteristics of the ideal-types posited by Kohlberg, Rest and other moral development theorists.

7.1.A.4. Low Gainers' Themes

Themes found in open and axial coding of Low Gainers' pre-essays were very illuminating. These three essays shared many similar themes with both Gainers and Decliners. Low Gainers' pre-essays exhibited some of the backward-looking quality of the Decliners, with Laura spending a fair amount of her essay reflecting on her father's unhappiness in his job. She reports that he is considering a career change that would bring him back to a job he had wanted while in college, more than 25 years ago. She notes that "regardless of his monetary success he is unfulfilled," and that some jobs "drive the humanity out of an individual as they are forced to value capital gain over personal relationships." The two other Low Gainers mention family and friends as well. Luke claims that "in a basic sense what makes me happy is being around the people that I love," but he seems uncertain about the sufficiency of this, noting that in his best life he "would have all of these basic necessities but on top of that material items of really high value." Lucy claims that one cannot be happy without family and friends and the love that radiates in and between them. She insists that to live best includes "remembering that I am loved and have done fun things with those who love me," reiterating the insistence on ties to family, friends and home common among Decliners. While Low Gainers often mentioned the need to search for purpose and pursue passion as Gainers had done, the

tones of their pre-essays were more confident, like Decliner essays. They posed no questions within their pre-essays, a consistent hallmark of the Gainer group. Low Gainers did acknowledge that finding happiness and purpose is a challenging search, a journey and takes effort, similar to Gainers' pre-essays. Laura goes so far as to claim that "it is each person's duty as a member of the global society to discover their [unique] talent and act upon it," possibly alluding to the uncertainty that her father's career change might present to the family.

Uncertainty and some degree of fear are marks of Low Gainers' pre-essays that are shared with Gainers. Luke writes at length about his present feeling of contentment and comfort, being "self-satisfied at this moment...I have no serious worries...[but] knowing that this could change in any instant." Laura has a similar nagging feeling of the temporary nature of these assertions, noting that "[p]eople could be more fulfilled and live without the regret my father lives with if they listen to their purpose, rather than to their wallets, no matter how difficult it is to do so." But for Gainers, that uncertainty resulted in a mandate to openness, whereas in Low Gainers, there is an air of sure opinion as seen regularly in Decliners. Also similar to Decliners, the Low Gainers seemed to crave balance. Lucy calls for balance several times in her pre-essay. She claims that we must balance taking care of oneself while trying to help others, "learning to balance selfishness with selflessness." Indeed, this theme is so important that Lucy mentions it six times in her last two paragraphs alone.

In post-essays, Low Gainers once again bring together aspects of both Gainers and Decliners in seemingly equal measure. Like the Gainers, each subject within the Low Gainer group focuses on finding a best way to live, or finding happiness as a *challenge* in

the face of persistent suffering that is part of life. Luke's early insistence in his pre-essay that his life of comfort is close to the best life, now acknowledges that "temporary happiness is easy to come by...[but] brings both happiness and sadness" eventually. He wonders now at the end of the year if happiness is even "possibly unachievable," and he concludes that "we as humankind simply will never live this way at the same time." He laments that suffering is such that at any given time, someone or some group will be unhappy, but wonders if we are "morally responsibility for at least attempting to make other's lives better." It's not surprising when reading this essay that Luke's rubric ratings were substantially higher than the other two Low Gainers and that his DIT N2 scores moved him from the lowest quartile in Pre-tests to the third quartile in post-tests.

Laura concurs with Luke's sense (and Gainers generally) that life must include facing struggle and suffering. She claims we must live in community "for better or worse" and that "it is in being exposed to corruption [and corrupt people] that one discovers the strength of his character." She highlights the life and writing of Viktor Frankl and admires his conclusion that love proves "that the good outweighs the bad...and gives one's life meaning." She goes on to claim that one must "trust in the strength of the absurd, in my case in the characters of relative strangers" to be happy and to live a full life, "even though it will inevitably lead to disappointment and incidences of heartbreak." Laura relies on several political philosophers to build her basic communitarian position, including Aristotle as well as modern philosophers Hobbes and Rousseau. In an important way, Laura is trying to reconcile what she sees as a righteous human search for community and meaning with the intrinsic suffering and heartache of life. But it still seems to her like an act of absurdity.

Though Lucy gained in her post-test N2 score, moving from the second to the third N2 score quartile, her essay shared much with Decliner post-essays. She tended to move to high abstraction, as many of in the Decliner group did, relying on a rather abstract notion of *authenticity* as the key to the best life. Interestingly, this same theme of authenticity was used as a central point by Decliner Diane. Though Diane and Lucy were from different classes they came to extraordinarily similar conclusions about the need to become “an authentic self” (more so even than becoming authentic, they both insist on becoming *this sort of self*, or performing this type of authentic personhood). Both focus on the constant *conflictual* state of life and conclude that becoming authentic is the only way to diminish these conflicts. Lucy points to the conflicts “between a person and the public, the second between the individual and the world, and the last between man and himself” and decides that we must at least try to resolve the conflicts within ourselves. But the example she ends up using, after a high level of abstraction about life and the need to become an “authentic self” is this: “for example, if an individual does not want to go to a party but goes anyway because he feels pressured, he is going to experience an internal tension. Instead, this person should obey his will and make other plans.” This, for Lucy, is what authenticity might be about. She mistakes the existentialist exhortation to authenticity for voluntarism or the determinacy of the will:

[i]f someone wants to dedicate his life to the service of others, it would be authentic of him to do so. If another wants to become a successful businessman and not give to others, it would be authentic to do this as well. Whichever way an individual choses to live his life, he should do it because he wants to.

Her desire to prioritize the individual's will results in a problem of instrumental relativism that Lucy can't quite work out. She goes on to conclude that "anyone who says they are perfectly happy is lying....All we can do as humans is to attempt to make the most out of what we are given. However, it is not easy. You may have to tell your friend that yes, her top is in fact ugly." These are not the high moral and globally aware concerns of the High to High Scorers' post-essays, or even the reiteration of how challenging it is to face the inherent suffering of life. Low Gainers' reflections remain mostly oriented to concrete and particular issues within their own limited horizons. Other people largely remain "other" and as such are quite outside the orbit of these Low Gainers' concerns. That Lucy cannot conceive of a deeper moral issue for her final paper's conclusion than how to tell a friend that her top is ugly reminds us of why her DIT scores remain low despite gains.

7.2 ESSAY ANALYSIS: THEMES/MOTIFS DETECTED IN SELECTIVE CODING

Selective coding offered a robust view of central motifs found across groups. Open coding provided a wealth of insights into themes in the writing of particular subjects and axial coding allowed the primary researcher to analyze general themes within each group. In order to gain adequate an adequate view *across groups*, however, selective coding sought a more holistic or bird's-eye view of patterns identified within groups for the purpose of parsing themes from the subsample as a whole. The selective coding process in this case entailed thoroughly reviewing the axial codes of each group and condensing these codes to several main core themes or central ideas that emerged in each group. The following analysis resulted from a comparison of this selective coding process.

7.2.1 Moral Development Themes across Groups

Analysis of student writing across the four student groups offers many insights into the moral reasoning development of first year college students. The sheer variety of characteristics exhibited across groups reveals the range of student moral development found among students enrolled in the type of liberal education program at the heart of this assessment. Findings from across-group analysis allows us to consider how student work may signal a student's position along that developmental range and may help educators become better at reading and responding to those signals. Within across-group analysis, the essays of Gainers, Decliners and Low Gainers were the easiest to compare and contrast. Themes found in High to High Scorers's essays were quite different from the other groups and yet those differences offer insights into refining programmatic objectives with an eye to increasing student moral development generally.

Juxtaposing Gainers and Decliners provided important areas of contrast, with Low Gainers sharing attributes of both of these groups. One of the most striking ways in which Gainers and Decliners differ is in the format of their pre-course essays. Gainers consistently posed their answers in terms of questions while Decliners offered very sure and solid opinions about the essay question. Indeed, in post-essays Decliners seemed hesitant to critique their own pre-essays and in some cases reiterated their first answers or indicated that they were simply adjusting their original ideas. Gainers' openness, on the other hand, was ratified not only by their claims that *one should be open* to questions and different perspectives, but also by the very use of questions throughout their essays. Gainers also regularly reported that adequately answering this question of the best way to live was a *challenge*. All Gainer essays referred to this as a *pursuit*, a *process*, a *search*

and all Gainers noted the high degree of difficulty of this pursuit. In this sense, Gainers present a very forward thinking posture and they see this pursuit as an active and demanding process. Many Gainers write about needing to leave behind old ideas and ways of thinking and being anxious about moving forward into unknown territory. The questions used in their essays exhibit how poised they are for development. Decliners' essays exhibit something quite contrary to this. Their essays are marked by a backward-looking quality. Most mention home, family and friends specifically and spend a fair amount of these short essays affirming how important these close circles are, referencing others outside their circles only in conceptual terms, as if the "world out there" is still only a vague idea. In some sense, Decliners reflect much more on where they have come from rather than where they are going. They focus on the need for *balance*, which no other subjects highlighted with the exception of one Low Gainer, Lucy. Interestingly, Lucy's post-essay was the only subject from the essay subsample to join Decliner David in logging a negative change score on the moral development rubric. Decliners' need for balance and safety seems to be an expression of a general anxiety of moving toward a future they perceive as demanding. While they express a desire for freedom (presumably associated with adulthood) and are deeply suspicious that social, ethical and political structures will limit those freedoms, their yearning for safety and balance reveals a basic fear of surrendering the security of childhood.

In post-essays, Decliners often move to high abstraction along ethical, personal and religious themes. At the end of the course many Decliners mention difficult personal challenges such as the death of a family member or a breakup with a romantic partner, pointing to the sort of circumstances that may impact or interrupt development. Unlike

Gainers, they do not see finding a happy or meaningful life as a difficult task or as a pursuit or struggle. They acknowledge the need to find one's purpose but this seems to them to be a discovery that will *simply happen*, while Gainers see this as a challenging task –something *to be done* and pursued. The question of personal ethical agency emerges here as an important distinction across groups. Gainers seem to be working out the connection of theory and real world issues as it relates to them. They scored much higher on the moral development rubric item, “Theory and real world connection” than Decliners or Low Gainers, and their post-essays advert to these concerns. However, they did not mention or reflect on specific theoretical models from course material that might help in this endeavor. They intimated concern for the connection of theory and personal ethical agency to real world problems, but ultimately chose personal ethics over larger construals of ethics or morality by the end of the course. Decliners similarly expressed concern about how things in the world ought to run but made no attempt to connect their own ethical agency to problems of the world or issues outside their own spheres. For them, the centrality of self was expressed in terms of personal interests, not personal agency in the world.

Not surprisingly, Low Gainer essays share qualities with both the Decliner and Gainer groups. Low Gainers see finding a happy or best life as a challenge or a difficult search, much like Gainers did in both pre- and post-essays. In pre-essays they posed this pursuit as active –something one must strive for, achieve, and discover –as Gainer pre-essays had. However, the general hesitation of the Gainer group and that group's propensity for posing numerous questions within their reflections was missing from this Low Gainer group. Instead, the Low Gainers expressed themselves in the same format as

Decliners, offering their answers in the sort of opinion format as opposed to a questioning format. Still, these Low Gainers also were more tentative about their answers than Decliners, acknowledging that their answers might need adjustment. They seemed anxious about what they didn't know but showed no indication of being ready to move toward larger social or moral issues. Echoing Perry's Multiplicity stage, these students evinced no desire to raise truly challenging or even possibly irreconcilable questions beyond the scope of their own circles, and yet they place great trust in their own instincts in seeking good and meaningful lives. Low Gainers also shared with Decliners a focus on home, family and friends though with less specificity. Like Decliners, Low Gainers expressed the need for balance, especially in regards to balancing self-interest and selflessness. This seemed to point to a basic tension in the Low Gainer essays between wanting to take the welfare of others into account but fearing that it would demand too much of them to do so. In post-essays, Low Gainers agreed that struggle and suffering are part of the search for a meaningful and happy life. They do not engage in the movement toward high abstraction found in Decliner post-essays, nor do they move to the level of questioning that is upheld and reiterated in Gainer post-essays. Rather, Low Gainer post-essays conclude largely that suffering is just a part of life that must be admitted: that heartache and struggle are inevitable, no life is perfect, and conflict is a necessary facet of community. But all Low Gainers state in post-essays that political, ethical and religious institutions are necessary, even though they will never be perfect. Low Gainers seem to be in the position of recognizing the limits of dualism and personal interest. They *seem* ready to admit that workable systems are needed for the advancement of all—even those outside “in-groups.” And yet, Low Gainers' pre- and post-essays exhibit a fairly incessant

circling back to themselves, a tendency that impedes a sustained consideration of the larger social demands of communal and cultural contexts.

As stated above, Decliner essays exhibited a great deal of backward-looking reflection. Specific family members and high school friends, as well as lessons learned in high school (both academic and personal insights) were prominent throughout these essays. Balance and safety were highlighted by these students but were completely absent in Gainer and High to High Scorer essays. Themes of physical health, basic human needs and aspects of homelife emerged regularly in Decliner essays but were absent in the essays of other subjects, with a few slight exceptions among Low Gainers. Decliners displayed a high level of abstraction in post-essays, perhaps as a mode of coping with challenges posed to their dualist viewpoints from events in their personal lives, coursework or general maturation. This movement to high abstraction suggests not a forward motion to a deeper consideration of larger communal or cultural contexts, but rather a retreat from those considerations to some other, more abstract dualism. For instance, Diane's post-essay focuses almost entirely on the demand for personal authenticity and pure freedom above all else, even though she claims the very idea of these is inherently absurd and cannot be reconciled in coherent social orders. This sort of retreating from the basic question of a best, meaningful or good life is a regular theme among Decliners and it communicates their sense of futility in the face of coordinating complex social demands. Decliners' preference for existentialist themes and texts from the course echoes this attitude of futility and the subsequent desire to retreat to the self as the starting point of any reflection on meaningful living.

This sense of the futility in finding a truly adequate answer to the question of the best life was also found among High to High Scorers' post-essays but here it is found in a new context. These high scorers also referenced existentialist themes as Decliners had, but posed Existentialism's insistence on the centrality of the individual as a necessary supplement to their very pronounced preference for large ethical and political considerations. Their doubts about the ever sufficiently answering the question of a best way to live were explicitly linked to a fear that individuals may get lost in the necessary process of establishing solid and progressive social, ethical and political systems. The futility expressed in these essays suggests something akin to Perry's advanced stage relativism, which allows for commitment to communal structures and systems. It is not surprising that High to High Scorers' pre- and post-essays offer a wide array of strong appeals for adequate social orders and political structures that ensure justice for all, with special emphasis on the marginalized. Essays from this group also highlight religious themes that coincide with these political goals, linking these social concerns to a holistic view of the person and society. High to High Scorers are careful to keep coming back to the concrete needs of individuals even as they make fairly grand claims about the need to give greatest attention to large-scale social, ethical and political structures. In pre-essays, High to High Scorers were already making clear statements about how closely linked individuals and society are in the pursuit of authentic and meaningful happiness. In post-essays, each of the High to High Scorers emphasized that the best way to live *must* involve active and simultaneous concern for individuals and communities. This is exactly the kind of macro-morality that Rest identified in his Postconventional Schema: an ability to fluidly move between the demands of macro-morality and micro-moral

concerns. That is, from the neo-Kohlbergian point of view, advancement in moral reasoning is evinced by one's capacity to hold in complex tension the need for large, working ethical and political systems and the exigencies of individual living and concrete realities. The High to High Scorers of this study are clearly excelling in reconciling this tension and finding nuanced and complex ways to approach issues of macro- and micro-morality, as shown both in their DIT N2 scores and in their writing.

7.2.2 Course-Related Themes across Groups

Rubric analysis of course-related items captured an overview of student engagement with the course materials across and within the different groups, as reflected in post-essays. However, axial and selective coding revealed interesting nuances regarding *how* students made use of course texts, ideas and class assignments that were not readily apparent in rubric analyses. Subjects' treatments of Aristotelian and ancient Greek ethical thought is an illuminating example. All groups included at least one subject who mentioned Aristotle, using his ideas on personal moral agency, friendship, or justice. Aristotle was one of the *commonly mentioned* figures in post-essays but there was wide variation in how the groups utilized this thinker's practical approach to ethics. Gainers and High to High Scorers demonstrated high engagement with Aristotelian thought, with many of these students adding Socrates and/or Plato in their reflections on practical ethics. Decliners and Low Gainers *mentioned* Aristotle's ethical ideas with some frequency but showed less engagement with those ideas overall. Most interesting, however, was how the groups differed in *juxtaposing* thinkers and ideas.

While only two Gainers' post-essays featured the ideas of Aristotle, all four Gainers highlighted his or other works such as that of Socrates and Jesus in order to

articulate the hard work and sacrifice that a truly happy life would entail. When Gainers did refer to Aristotle, they combined his practical ideas with the works and words of Viktor Frankl, Martin Luther King, Jr., Socrates and Jesus, all figures who suffered in the pursuit of a good, ethical life. Gainers seem to present these figures aspirationally, as models of how a good life might actually look when things inevitably get challenging. For them, Aristotle's approach to a life of happiness and goodness offers a theoretical structure that also practically addresses the fragility of human living. In the case of a third Gainer, Socrates' teachings were posited as a helpful guide and reiterate the importance of openness, titling her paper, "The Importance of Questions According to Socrates, Plato, Martin Luther King, Jr. and Me." Combining Greek thought on the development of virtue and ethics with figures whose work and lives point to the great demands of the ethical life is a common and central theme found among Gainers. This echoes general themes found throughout Gainer essays of their awareness of and openness to the challenges of moving forward to ethical engagement in the world.

High to High Scorers share this engagement with ancient Greek ethics. However, they combine Greek philosophical thought with contemporary, global issues, rather than focusing this engagement on questions or figures of the past. High to High Scorers were much more apt to bring Aristotle into dialogue with documentaries they had viewed, questions about the inherent fairness of modern economic systems, contemporary social justice issues, or how to reconcile the rights of individuals and social structures. This tendency affirms the moral development rubric analysis' finding of High to High Scorers' advanced capacity to hold macro- and micro-moral questions in tension and their aptitude for comprehending the lived implications of theoretical understanding.

Contrastingly, the two Decliners who mentioned Aristotle or ancient Greek thought in post-essays used these ideas quite differently than Gainers and High to High Scorers. In one case, Aristotle's thought was presented in highly abstract and theoretical terms and in another, only Aristotle's theory of friendship was mentioned in the context of the subject's assertion that living a good life "require[s] relationships with other people." This, the author suggests "is a very realistic viewpoint, that people in society struggle between altruism and serving only themselves." Decliners concern with the self seems to block their attempts to deeply engage with Greek notions of a practical ethics of agency and its implications for contemporary issues.

Finally, Low Gainers were even less apt to mention Aristotle or engage in ancient Greek ethical thought, but did mention Viktor Frankl's reflections on suffering or "the suffering of the Jews in the Bible," reminiscent of Gainer post-essays. While Low Gainers saw courage and optimism in the face of suffering as admirable, they gave their assent to it begrudgingly. As one subject put it, "[i]f Viktor Frankl, who lived through the worst evil that community living has produced, can still say that living with others is worth it...then it is safe to say that the good outweighs the bad." But this, she concludes, includes our ability to "trust in the strength of the absurd."

Coursework is clearly perceived and processed differently among these groups. A student's ability to apprehend the relationships between course materials, her own life and the larger issues and problems of her times is evidently connected to her moral development. Whether one's ability to make those connections is a cause or effect of moral development is unclear but it seems apparent that a student's reception of course material and her capacity to grasp the connection of coursework to "real life" tells us a lot

about her moral development generally and how that development might present itself to educators.

7.3 ESSAY ANALYSES: LIMITATIONS OF FINDINGS

There are several important limitations to consider regarding the findings from the present study's analysis of student writing and any conclusions that might be drawn from them. First, there is the matter of possible researcher bias. The subsample of students included in the qualitative component of the study was a convenience sample that included two classes out of sixteen classes included in the program at the heart of the study. One of these classes was taught by the primary researcher. Thus, of the 15 students whose essays were eventually selected for the writing analysis, 8 were students in the primary researcher's class. Though the researcher has taught the class for 16 years prior to this study, it is possible that she unintentionally taught the course in a way that implicitly or explicitly advantaged student performance on the DIT and student writing along rubric lines. Additionally, although names were removed from student writing and aliases were assigned, it is possible that the primary researcher's knowledge of the students and their general attitudes and behaviors would add confounding bias to the analysis process. This type of researcher bias is a very serious consideration, especially in qualitative research. However, the mixed-methods approach of the present study may ameliorate researcher bias to some degree, since DIT scores were used to determine student groups rather than teachers in the program or the primary researcher. Additionally, rubric analyses that were piloted with a second rater (who was not involved in the program being studied), were helpful not only in rating students on a variety of literature-based moral development and course-related items, but also aided the writing

analysis by highlighting aspects of essays that were important, rather than depending on the researcher's understanding or conceiving of student ideas. It is hoped that these aspects of the study mitigate some of the study's possible researcher-based bias.

Limitations of the study also include concerns about the students involved in the study's writing subsample. An ideal sample for the writing analysis would have included a random selection of students from all classes participating in the DIT, but unfortunately, only these two classes used both pre- and post-course essay assignment. The impact of teaching practice, demographic diversity within the class, student self-selection into the course and into these two classes specifically, may all confound the findings as well but were not measured in this study. However, particular qualities of the course, including the relatively small class size and full year format, offered students not only depth of contact with ideas and materials of the course but also offered a sustained and communal discussion of ethical, social, political and moral questions at a time when researchers have found young adults making great strides in moral reasoning development. As stated earlier in this work, research has demonstrated very clearly that college participation positively impacts moral reasoning development, with particular advancement in the first year of college. Moreover, research shows that this development does not reduce to intelligence or to a number of other factors one might assume underlies this sort of development. In order to get a better sense of how college student moral development actually happens and how specific college experiences influence that development, it is important to listen carefully to what students themselves report about their own ideas about how one best constructs and lives a meaningful and purposeful life.

7.4 ESSAY ANALYSES: CONCLUSION

Despite potential limitations, the qualitative component of the present study offers an important element to a thorough and robust examination of student moral reasoning development. While quantitative analysis via the DIT and rubric analyses offer generalizable data about students and their experiences in a college course or program, qualitative analysis uncovers facets of moral development that escape quantitative measures. How students perceive their own development and how capable they are in articulating *different and differently changing* aspects of moral development are imperative for gaining a full picture of advancement in this developmental area.

Having pre- and post-course essays in this case offered a unique view of a variety of cognitive and affective movements associated with moral reasoning development. The essays at the center of this part of the study showed substantial variation in both students' moral reasoning development at the beginning and end of the year-long course and the extent to which course-related materials impacted or interacted with that development. However, the writing analysis also contributed two important and unique facets of student moral development that are of interest to educators: first, the analysis demonstrates how differently students at various stages of development receive and process coursework and course materials; second, the analysis reveals important features of moral development that educators may easily identify to better instill, support and build upon moral reasoning advancement. Knowing what moral reasoning looks like in students who are advanced in this area and who gain greatly in their first year, as well as knowing what blocks development may provide educators with important considerations for designing courses, creating responsible and reasonable educational objectives and pedagogical strategies that actually correspond to student development.

Chapter 8:

Implications of the Study and Conclusion

8.1 REVIEW OF THE RESEARCH QUESTIONS AND DESIGN

Moral development is a notoriously elusive area of development research, simultaneously widely contested and sought in American higher education. In that arena, the advancement of moral development boasts a wide array of concerned stakeholders, from college professors and administrators to politicians, employers, students and parents. There is much at stake in how we address what moral development is, what it is not, whether and to what extent it can be measured, and what educators can do to set the stage for it. These issues are particularly important for those who design and implement liberal education programs, since moral development has long been claimed as a goal of that educational paradigm. As public debates intensify over the cost, value and purpose of higher education, educators in the liberal education tradition have good reason to defend moral development as a key objective but must demonstrate its ability to deliver on this claim. Evidence-based assessment is a critical piece of this endeavor. Against the backdrop of public and political demands for accountability accompanied by increasing shifts toward vocational and pre-professional higher education, liberal education needs to prove its relevance and value. To that end, liberal education, which has not historically been oriented toward research and evidence-based design and evaluation, must begin to fruitfully assess the objectives it claims to uniquely advance. Those aims extend far beyond purely intellectual pursuits. The persistence of moral and ethical development in American college and university mission statements testifies to its enduring presence. But an important question remains: can the moral aims of liberal arts education be adequately

measured and if so, what can we discover about the sorts of experiences that motivate, promote and sustain moral growth?

The present study contributes to addressing these questions by examining data on one aspect of moral development –that of moral reasoning development (described in detail in Chapter 2)– among first year college students enrolled in a liberal education program at a research university with a robust liberal arts core curriculum. Moral development literature suggests that the period between adolescence and young adulthood is a very productive time for moral reasoning growth and that this area of development is highly active in the first year of college, especially in the context of residential four-year colleges (this is the case even when a number of related factors are controlled, such as intelligence, college readiness, race, gender, and so forth). As researchers Patricia King and Mathew Mayhew put it, “intentionally or unintentionally, moral development is an outcome of higher education” (2002, p. 249). These researchers have joined others in exploring the impact of various aspects of liberal education –course materials, pedagogical strategies, classroom sizes and times, teaching styles, etc.– to uncover aspects of liberal education that might be uniquely effective in advancing student moral reasoning. However, research is only just beginning to help us understand precisely how and why that development is happening and which particular aspects of college are most influential in this advancement.

This study takes an important step toward understanding the association of moral reasoning development and liberal education coursework that seeks to set the conditions for that development. Results of the study illuminate important features of student moral reasoning that should guide the planning, implementation and evaluation of courses that

anticipate moral development objectives. The study's findings indicate wide variation of moral reasoning development and developmental gains in a group of first year college students during a year-long liberal education course designed to engage students in questions of justice, ethics, virtue and the good life through a rigorous study of the philosophical and theological foundations of western thought. This variation in student development interacts in important ways with students' ability to receive, engage and make meaning of what they encounter in the educational process. Implications of these results fall into two main categories: first, the study offers a wealth of information about the moral reasoning development of a group of first year college students and how their levels of development and gains over time interact with liberal education coursework; second, the results provide insights into effective assessment practices for liberal education courses and programs. This second set of implications raises a number of important questions for practitioners of liberal education as they begin to take seriously the charge to evaluate programmatic objectives using data-driven methods that have not traditionally sat comfortably within that educational paradigm.

8.2 SUMMARY OF THE STUDY FINDINGS

The study focused on three central research questions. The first research question, whether this course resulted in moral development growth, is answered affirmatively. The second inquiry investigated how students perceive the moral dimensions of the course and the third inquiry explored the relation of those course dimensions to student moral reasoning development. Despite variation found in student perception of the course, triangulated quantitative and qualitative data aids us in gaining a full picture of students' engagement with aspects of the course and the impact of that engagement on

moral development. Distinguishing subsample groups allowed an exploration of the attributes, concerns, opportunities and impediments associated with the moral growth of students at different stages of development. The following section will explain the study's affirmation of the first research question and will address results related to the second and third questions. Interpretations of the results will be presented in section 8.3 and implications of the findings will be presented in sections 8.4 and 8.5.

Students in the study sample demonstrated tremendous growth over the year of the course as compared with national norms. Even though mean pre-test scores of this sample showed a very high starting point for this sample relative to researched national norms, mean *gains* of this sample were commensurate with typical gains reported over *four years* of college in aggregated research. While the quasi-experimental nature of the study cannot establish the cause of these sorts of gains, the study takes important steps in exploring which aspects of the educational experience might be influential in this development. Great variation was identified among the DIT scores of students in the course though mean N2 scores of the whole sample were found to be significantly higher than national norms for college students *in both pre- and post-tests*. In fact, the pre-test mean scores of this sample more closely match national figures for Master's degree students and mean post-test scores came in even higher, rivaling nationally normed scores of students in advanced graduate professional degree programs. It is apparent that a significant percentage of students in this course present with very high levels of moral reasoning compared to other students of the same age and educational level, and most students will make significant gains while enrolled in first year college courses. Additionally, students on the whole will present with a wide array of moral reasoning

capacities, from these very high levels to quite low levels of moral reasoning, and that development will change positively or in some cases negatively in widely divergent ways within the context of our courses and programs.

The type of variation found across the study sample in terms of DIT pre- and post-test N2 scores and findings from analyses of student writing is instructive. DIT outcomes showed a wide range of student development including low-level scorers, subjects who gained or declined precipitously in N2 scores, students who experienced gains from very low points along the DIT spectrum, and students whose scores were substantially higher than national norms and their classmates. Results showed that 12% of the sample posted very high N2 scores at both the start and end of the course, while 5% of the sample posted N2 scores well below national norms for students of that age and educational level. While the mean gains of the sample were notable, 27% of the sample made substantial gains of more than one standard deviation, with 11% of the sample making these sorts of gains from quite low starting points. The picture was not entirely rosy, however. Seventeen percent of the sample posted declines in N2 scores by the end of the course, though steep declines (a full standard deviation) were rare. Using four of these categories, High to High Scorers, Gainers, Low Gainers and Decliners, a helpful picture of moral development attributes and course engagement patterns emerged from student writing analyses within and across groups. Two moral development rubrics including 15 literature-based components of moral reasoning (such as “Prosocial behavior” and “Civic engagement attitude”) and 11 course related items (such as “Course content as interruption” and “Aristotelian ethics”), offered a close examination of which components of moral reasoning and which features of the course were actively

emphasized (or underemphasized) by students in their writing about living “the good life.” Variation within student writing was remarkable and is indicative of how essential qualitative analysis is in explaining more robustly what moral reasoning development data means. The study found high variation *across* these student groups and great consistency *within* groups regarding notable aspects of moral development in which students seemed most or least engaged. Those same patterns of variation and consistency were identified in student reporting of how coursework interacted with their evolving abilities to address questions with moral dimensions.

Students with very high moral reasoning scores displayed a capacity for pivoting fluidly between micro- and macro-moral concerns and, though concerned with the establishment of functioning systems and institutions, always tracked back to a care and concern for individuals even over and against societal demands. They showed greater preference for connecting real world issues to the ethical and political theory than their classmates and the scope of their moral, ethical and political concerns sought wider and more diverse contexts than students at lower levels of development. These high scorers consistently brought textual analysis of justice issues into dialogue with contemporary issues and expressed openness to diverse and diverging opinions on the good life. High Scorers consistently made connections between class materials and other ideas, texts, lectures, documentaries and activities from other parts of their lives. Other students, whose N2 post-scores exhibited great gains but who did not reach these highest moral reasoning elevations, shared with high scorers a proclivity for combining theory and real world concerns. Their essays added their sense of how hard they perceived this pursuit to be. Gainers consistently viewed the search for the good life as a struggle but did not shy

away from the challenge. The most startlingly common feature of Gainers' writings was an almost incessant use of questions, mostly directed to themselves, about what a life of justice and ethics might entail. It is notable that no student from outside the gainer group posed a single question in essays, while this was a common feature in Gainer essays. In post-essays this preference for questions was combined with an appreciation of thinkers who exemplified an "examined life" and engaged in dogged or even self-sacrificial pursuits of justice.

At the other end of the moral development spectrum, students posting declining N2 scores took a distinctly different approach to reflecting on the good life. They were overly and overtly confident in their reflections, offering no questions and little acknowledgment that the question of a good life posed any difficulty. Writing from this group exhibited a backward-looking tendency, as though home and past friendships and experiences had taught them enough to sufficiently address the question. Decliners and Gainers privileged principled thinking and apprehended personal agency in the welfare of others, but Decliners scored lowest of all groups in integrating diverse perspectives and acknowledging the limitations of their own understandings and contexts. Comprehensive assessment like this can pinpoint significant differences in the ways that one type of student makes meaning of moral concerns as opposed to another. Students who showed declines in moral reasoning and students who gained from very low pre-test positions share a basic preference for existentialist and Nietzschean themes within course material as well as a relative distaste for theory. Low Gainers' share with Gainers an apprehension of the challenge of pursuing ideas about the good life, but their tendency to prioritize

individual experience while eschewing theory reveals how much like their Decliner classmates these Low Gainers are with respect to reception of course materials.

8.3 INTERPRETATION OF THE STUDY FINDINGS

8.3.A Interpretation of DIT Findings

Findings in this study strongly support moral reasoning research claims. However, adequate interpretation of these findings must address some of the unexpected aspects of the results. DIT *scores and gains* found in the study were both substantially higher than national norms (Rest, Narvaez, Bebeau & Thoma, 1999b, 2000), which raises questions about the validity of the sample and the influence of confounding variables. The first set of concerns is associated with particular attributes of the sample and course investigated in the study. The course at the center of the study is offered at a highly selective university. Due to this selectivity, we assume that students enrolled in this university would have significantly higher intellectual ability compared with national trends and since intelligence is positively associated with the development of moral reasoning it makes sense that scores of students at a highly selective college would post higher DIT N2 scores (King & Kitchener, 1994; King & Mayhew, 2004). Moreover, the observed course is known for its rigor and enrollment in the course is limited so the course likely attracts students with high cognitive motivation which is also found to be associated with moral reasoning growth (Mayhew, Seifert & Pascarella, 2010; Mayhew, Wolniak & Pascarella, 2008). One might argue that the high moral reasoning scores and gains found in the study reflect advanced intellectual ability, cognitive complexity or advanced cognitive motivation. Interestingly, the study sample, which included students from across the four areas of the university (including Schools of Nursing, Education, Arts and

Sciences, and Management) posted mean N2 pre-test scores that were more than 5 points higher than mean pre-test scores of students from one area of that same university (the School of Management), to which students are generally admitted with higher SAT/ACT scores (Sullivan, 2011). In other words, while we might assume that higher intellectual or cognitive capacity as measured via the SAT would be associated with higher moral reasoning scores, it did not bear out in this comparison (admitting that these groups cannot be robustly compared). Furthermore, DIT research (particularly in the past decade) corroborates Rest's assertion that moral reasoning is not simply cognitive development but is a social cognitive development that involves different kinds of interactions between cognition, affect and tacit construals of morality (King & Mayhew, 2002; Mayhew & King, 2008; Mayhew, Seifert & Pascarella, 2010; Mayhew, Wolniak & Pascarella, 2008; Narvaez, 1999, 2001; Rest et al., 1999b; Thoma, 2002). Indeed, King and Mayhew's meta-analysis of 172 studies of moral reasoning development finds that while high levels of moral reasoning are associated with complex, higher-order cognitive activity, cognition is a necessary but not sufficient condition for moral reasoning growth (2002). Results from the study also pay primary attention to the differences *among the students in the sample* rather than focusing on a comparison of these students with a control group or national norms and trends. The triangulation model of the study offers a means to view the influence of this course on the moral growth of these students, rather than looking to show evidence that these students have more or less moral reasoning capacity than others outside the study. Finally, it is important to note that the study seeks primarily to address the question, Can liberal education be assessed? by uncovering how students across the range of moral reasoning development enter, change and leave their

first year of college (with respect to that development) and what we can discern from their writing about the ways that coursework interacted with that development.

Results of the study highlight the tremendous growth that most students experience in the first year of college in terms of their increasingly complex and nuanced abilities to address moral issues. They also remind us that development in this area is much more than simply a matter of growing up (King & Kitchener, 1994, 2004; King & Mayhew, 2002). Findings from this study give a broad picture of moral reasoning gains across the spectrum of scorers and as such, tell a story of the wide variation in development and developmental gains that we find among first year college students with respect to moral reasoning. Gains may be found at all levels of development and those gains look quite different at different levels of development. This reinforces Pascarella's suggestion that great variation among college students' moral reasoning development may explain why the cumulative effect of college experiences is pronounced and large as opposed to that of a single or particular experience (1997). Examining two different types of gainers is an important step in uncovering which aspects of moral development and course engagement tend to be associated with developmental gains while others are closely associated moral reasoning level consolidation (Mayhew, Seifert & Pascarella, 2012; Thoma & Rest, 1999). Results show, for instance, that Low Gainers exhibit course-related preferences that are more like those of Decliners than those of other Gainers. On the other hand, Low Gainers are more similar to Gainers in seeing the quest for a good life as a challenge, search or pursuit and lack the overt self-assurance of the Decliners. Results related to students whose scores declined over the year provide a fascinating snapshot of the sort of retrenchment that some developmental theories propose. One

might argue that this negative change contravenes neo-Kohlbergian research that posits upwardly directed development. Rest's schema theory suggests that once subjects move to higher modes of moral reasoning, they do not lose these capacities (Rest et al., 1999b). However, the schema theory argues that a subject is at any given time demonstrating preference for one schema over another and in transition phases may appear to move back and forth between schemas. Periods of transition and consolidation may account for minor gains and losses. It is notable that the study's original group parameters had to be adjusted to find adequate numbers of students whose DIT scores declined precipitously. There simply were not enough students in the sample whose scores showed significant post-test losses as the original group parameters speculated, though there was an adequate group of students to consider whose scores had dropped somewhat. Findings related to subjects who lost ground reveal that most of these students came from the lowest and second lowest quartiles of pre-test scores and none came from the top pre-test quartile. In other words, most students who did decline in DIT scores did so at lower levels, where personal interest schema concerns might reemerge and restrain development before consolidation in the Maintaining Norms or Postconventional schemas occurs. Moral disengagement and developmental retrenchment at significant transitional points have been suggested by other developmental models and may be supported by these findings (Bandura, 2002; Bandura, Barbaranelli, Caprara & Pastorelli, 1996; Perry, 1999). Still, the characteristics of Decliner activity in this study do not contradict DIT research claims that subjects tend to not slide back once they have achieved *higher levels* of moral reasoning.

The quantitative findings of this study are in and of themselves quite robust and helpful. Data from the study offer insights into the relative presence of students at very high or very low levels of moral development, as well as students who are apt to make significant gains or whose moral reasoning capacity might regress. Using DIT data to establish categories of high and low scorers and high and low gainers allows a coherent account of findings despite the great variation found therein. These groupings offer a way to assess student positionality along the range of moral development and to think more cogently about how to best address the needs of students at those different points in their development. In sum, quantitative data show that moral reasoning development is not a simple and straightforward trajectory. It is a multifaceted process, the parts of which often move in piecemeal and fragmented ways. Results of this study show that not all aspects of moral reasoning are created equal, so to speak. That is, pieces of moral reasoning seem to emerge separately and are utilized to different extents by different subjects. It is when the pieces begin to coalesce that we see the kind of *preference for* advanced schemes over lower schemes (as opposed to a Kohlbergian hard stage model), described by Rest and identified in the DIT (1999b). In the process of moral or ethical meaning-making, the various components of moral reasoning identified in the moral development rubric of this study (and described in detail in Chapter 4) are activated. As students advance along these various moral reasoning components, they become more adept in resolving questions that are insufficiently answered by lower order moral thinking. The data from this study suggest that the extent to which students make use of the components of moral reasoning affects their ability to make sense of moral dilemmas, moral questions and the demands of living a moral or ethical life.

8.3.B Interpretation of Writing Analyses

Analysis of student writing via two rubrics and several rounds of essay coding fleshed out quantitative results from across- and within-group developmental differences. Insights from these analyses take important steps in apprehending and understanding various components of moral reasoning and their relative activation and mobilization in educational contexts. Examining students' own articulations about living "a good life" and analyzing their engagement with perceived impactful aspects of the course allows us to scrutinize specific aspects of moral reasoning to uncover which components of moral reasoning are active at high or low levels of development and at high or low levels of developmental change. This investigation also allows a view of aspects of moral reasoning that may be sluggish or dormant at lower levels of development or growth.

All of the students from the study's essay subsample prioritized living according to universalizable and widely held principles, and most exhibited a relatively advanced capacity to exclude simplistic ways of thinking. But only students at very high levels of moral reasoning development demonstrated engagement with differing perspectives and diverse views on how to live a good life. The pursuit of principled living appeals to most students *in principle*, so to speak, but the practical reality of doing so demands that subjects gain a broader and more adequate apprehension of "the other" beyond one's in-group. Results of this study imply that the central difference between those at higher or lower levels of development and change lies in the ability to expand circles of concern. Subjects at lower levels of development and developmental change adverted to macro-moral issues but only in abstract ways and they seemed unable to think about how their own lives were connected to these larger concerns because they could not view those

concerns from a point of view other than their own. Additionally, they presented their care and concern for people in their own circles as evidence of moral concern for others. In neo-Kohlbergian terms, this tendency to privilege in-group concerns and ignore “the other” is a main impediment to Postconventional thinking (Rest et al., 1999b). Students at very high levels of development, on the other hand, displayed expansive and highly inclusive notions of circles of concern. Their high engagement in prosocial attitudes and prosocial behaviors such as participation in service trips, work for political causes, connecting course material to concrete, global suffering, and so forth, give further evidence of the importance of broad construals of “the other” at higher levels of moral development. An interesting conclusion may be drawn about a particular aspect of moral reasoning that research has begun to explore. Nascent research has begun on the effect of cognitive motivation on moral reasoning. This feature of development refers to a willingness on the part of an individual to engage in effortful thinking and has been associated with moral reasoning development (Cacioppo, Perry & Kao, 1984; King & Mayhew, 2002; Mayhew & King, 2008; Mayhew, Seifert & Pascarella, 2010; Mayhew, Wolniak & Pascarella, 2008). This study’s results suggest that a similar attribute of moral reasoning might be necessary to growth: a willingness on the part of an individual to engage in effortful consideration of “the other.” Researcher Matthew Mayhew has examined the effects of experiences of diversity on students’ moral development, including diverse peer interaction and exposure to diverse thought (Mayhew, 2004; Mayhew & Engberg, 2010; Mayhew & Fernandez, 2007; Mayhew & King, 2008; Mayhew, Wolniak, Pascarella, 2008) but how willing a student may be to be effortful in the specific task of expanding one’s conception of “the other” has not been studied.

Essay analysis revealed a nuanced view of *how* students approached, received and made use of course texts, ideas and assignments throughout the year. Indeed, the divergent approaches used in addressing the essay question discloses a lot about students' positions along the range of moral development and their capacity to engage with course materials in effective ways. An openness to questions and a Socratic willingness to admit what they did not know was an important feature of students who posted substantial moral gains. A corollary willingness to take on the "big" questions of life, even at personal cost, was associated with growth. Gainers seemed to find great consolation in reflecting on thinkers who made personal sacrifices to live according to principles, regularly juxtaposing theory with these examples. These same course activities were recognized by students at lower levels of development but were conceived abstractly and not as examples that point to concrete ways of living. Students at lower developmental levels seem to process course material in such a way as to hold ideas and examples at arm's length since they are seemingly unready or unwilling to integrate these fully.

High scoring students and high gainers shared a preference for ethical theory and ethical political theory and combined these with strong displays of theory and real world connections. Decliners shared this preference for ethical political theory but were strikingly disengaged with the connection of real world issues and theory. They exhibited greater engagement with existentialist themes, which allowed them to focus on themselves and to shut down questions concerning those outside their own, limited horizons. The certainty and confidence with which Decliners approached their pre-essays, their persistent emphasis on balance and safety, and their backward-looking tendencies give ample evidence of the intricate and extensive personal and emotional growth that

must accompany moral development. Moreover, it is helpful to recall that several Decliners reported recent traumatic experiences, including the death of a parent and a difficult romantic break-up. Interestingly, only one other student in the study, a Low Gainer, mentioned the effects of a traumatic event (her father's job loss and subsequent long-term unemployment) and her pre- and post-test N2 scores were the lowest of that particular group. While we cannot control the exigencies of personal loss and trauma that may impede development, it is important to remember how impactful these kinds of experiences may be and how much they may thwart our efforts to foster moral development. Moral disengagement theory may offer insights into how and why these experiences short-circuit moral growth (Bandura, 2002; Bandura, Barbaranelli, Caprara & Pastorelli, 1996).

8.3.C Limitations of the Study

This study employed a mixed-method, longitudinal, theory-based approach to evaluate student moral development and course-related perceptions and employed a triangulation design to uncover the interaction of student moral growth and coursework (Saldaña, 2011; Stake, 1995). A well-researched measure of moral reasoning was combined with two rubrics developed by the researcher for student essay analysis in the triangulation of data. Findings from the study are fairly comprehensive but also admit of some important limitations. First, the study suffers from a lack of an control group, since all of the students in the study were exposed to the intervention investigated in the study. Thus, the comparison group used in the study included aggregated national norms, making it quite difficult to determine the accuracy of the results. The comparison of data was complicated by the high selectivity of both the university and the course in question,

both of which introduce the possibility of confounding variables associated with student self-selection, college readiness, cognitive motivation, and so forth. A feasible control group could include students who wanted to take the course but didn't get in and enrolled in one of the university's many sequenced, two-semester courses or for another year-long course. This would ameliorate the problem of self-selection since it could be assumed that these same students would have selected into the course if possible. As to other limitations of the study design, it would be impossible to randomize selection into the treatment without making major changes to the way the course is administered and populated. Since the course is known to be rigorous, many students would not want to be randomly placed in the course. Thus attempts to evaluate this particular program will almost inevitably be confounded by self-selection issues. In future assessment designs, an ideal control group would include a course outside the liberal education tradition, perhaps a pre-professional ethics course, though most of these would be one semester courses only. To resolve this issue, one might argue for measuring students' moral reasoning in this course at the end of only one semester but the course uses a full-year design and, as such, it would do the course a disservice to measure its goals at only the end of the first half of the course. Comparing the effect of this particular course against another, similar course is possible, but this would not address the central feature of the question regarding the impact of liberal education on moral reasoning.

A second drawback of this investigation is the relatively limited scope of the sample. The study examines data on one group of students in one program at one university and thus is very limited in its generalizability. Since the study employed a secondary analysis of course assessment materials, it was difficult to avoid this limitation.

In the future, program evaluators may want to expand the assessment to include further pertinent data and identify and include comparison groups from other universities or colleges. Exploring various attributes of different educational environments would open up several paths of inquiry that were not available in this study.

The inclusion of a limited array of data in this study is a third area that may weaken the study's findings. Other studies of moral reasoning development include many data points that are relevant to moral development including measures of intelligence, cognitive motivation, college readiness, classroom practices, extracurricular patterns, diverse peer experiences, class time and size, and so forth. Future attempts to gain assess the effects of this course might include research into other salient aspects like the ones mentioned here.

A final, possibly limiting factor of the study regards the positionality of the researcher, who was the instructor in one of the classes at the heart of the study. Issues related to this concern are addressed in detail in Chapter 4.

8.4 IMPLICATIONS FOR THE COURSE, PERSPECTIVES IN WESTERN CULTURE

Overall, the study concludes that participation in this liberal education course did positively influence the moral reasoning growth of the students involved in the study. Educators in this program would be well served by reflecting on the range of development and of developmental gains found in this study. They would also be well served by knowing what sort of moral development they should expect to find among their students, particularly as they conceive moral and ethical development in course and program objectives. The groups identified and described in this study provide a basic breakdown of the types of students populating our classes as well as salient attributes of

the moral development of students in these groups. This study proposes that highlighting major groups like these – students whose moral reasoning capacity is quite high or quite low and students who show great developmental gains or losses in this area– reveals how multidimensional and versatile course moral development goals should be. Since students come into our courses at such different developmental starting points, it makes sense to have a comprehensive and flexible set of goals in this area. What we seek to accomplish is not simply for all students to rise to the level of moral reasoning of graduate students or to achieve one particular developmental level, but to build on the capacities they already have when they enroll in our classes, to promote gains in the area of development we purport to foster and to aid students in maintaining the development they have achieved. A number of recommendations emerge from the study results.

While most faculty members are adept at assessing their students' intellectual ability, their grasp of central ideas and their ability to write coherently, it would certainly surprise most faculty members to discover that some of the first year students in their classes have the moral reasoning capacity of a first or second year graduate student. In light of these findings, faculty members should be reminded that meeting the needs of their top students should not be limited to intellectual challenges but should include moral and ethical challenges as well. Moreover, we must remember that while facets of moral development are closely associated with cognitive and intellectual development, it really is an area of development that demands coherent pedagogical attention in its own right. In dealing with students who display the attributes of very high levels of moral reasoning, educators should note that students themselves may very well model *for us* what works. Students at very high levels of moral reasoning demonstrate attributes that

we should pay attention to since these attributes serve to recommend pedagogical practices that may activate the kinds of attributes needed for growth.

Research has demonstrated the positive effects of different pedagogical strategies and experiential classroom practices such as perspective-taking, intergroup dialogue, guided inquiry, dilemma discussions, role taking, and active learning (Mayhew, 2004; Mayhew & Engberg, 2010; Mayhew & Fernandez, 2007; Mayhew & King, 2008; Mayhew, Wolniak & Pascarella, 2008; Mentkowski, 2000; Grunwald & Mayhew, 2008; Reiman, 2004; Skoe, 2010). Although there are no specific experiential extracurricular activities required in this course, faculty design and build into curriculum experiential components that facilitate these sorts of effective practices. Small group discussions, peer-to-peer interviews and written assignments are helpful in setting the conditions for students to encounter the lived, moral dimensions of the course and of course materials. A further step might include *requiring* students to draw parallels between this course and other experiences, classes, ideas and activities or allocating percentages of grades to these activities. The regular, focused discussions on the connection of course materials to current political events that most faculty make time for in this course capitalize on the habits of high level students to bridge theory to real world concerns. Assigning attendance at campus lectures or documentary viewings in group settings and asking students to identify the central moral question within the topic at hand is used in limited ways by faculty. This should be encouraged more regularly, since it clearly mobilize moral sensitivity, an important feature of moral development (Bebeau, 1987, 1993; Bebeau & Brabeck, 1987; Rest et al., 1999b) and serves to bring coursework into dialogue with contemporary concerns. This format also sets the conditions for students to

discuss global implications of issues with each other in order to activate the effect of positive diverse peer interaction (Mayhew, Wolniak & Pascarella, 2008). These topics may also be addressed in larger class discussions and grounded in ethical or textual analysis.

Within the context of enacting textual analyses of current event and experiences, best practices include presenting diverse interpretations and perspectives in class in an attempt to move dualists toward positions of relativism and to encourage students to build habits of perspective-taking and holding competing points of view simultaneously (Mayhew et al., 2008). These are practices in which high order thinkers will naturally engage. However, recent research suggests that “merely exposing students to diverse perspectives may not be enough to disrupt existing schema associated with egocentric frames for understanding justice, as many students may retreat from, rather than work through, the discomfort engendered by confronting unfamiliar perspectives” (Mayhew & Engberg, 2010; Mayhew, Siefert, Pascarella, Laird, Blaich, 2012). Investigations into this particular pedagogical strategy recommend developing models that teach students strategies for working through discomfort and confrontation that invite broadened frameworks and widened circles of concern. Capitalizing on students’ habits that are already regularly activated at various levels of development would be a good first step in developing those models. Mayhew suggests that presenting series of competing viewpoints within course content enacts integration that is central to the meaning-making processes of students. The content of this course lends itself well to this strategy, since the respected thinkers presented in the course present differing views, criticize each other’s work and encourage confrontation of ideas. Thus, educators in this course should

highlight and make regular use of this feature of the course in an attempt to help students avoid developmental retrenchment and retreat.

Results from this study affirm the efficacy and primacy of a rigorous practice of questioning one's assumptions and conventional knowledge. Findings from Gainer essays signaled that students on the precipice of making significant developmental gains engage in persistent and consequential questioning. This activity relates to the basic assumptions undergirding the DIT itself—that the dilemma format includes and instigates fruitful questions that trigger tacit moral understandings and expose those understandings as sufficient or insufficient to the demands of the question (Narvaez and Bock, 2002; Rest, 1973, 1999b). It is important then that we as educators model for students the basic premise that good questions are as important as good answers. Helping students grapple with and expand their questions by pointing out further pertinent facets of the line of inquiry and connecting these to other meaningful questions, rather than trying to quickly answer them, encourages the sort of openness that Gainers displayed in essays. Being persistent in the pursuit of good questions is apparently central to how students pursue making meaning of moral and ethical concerns.

Insights related to data from students whose N2 scores declined are instructive for thinking about how to address key pieces of moral development that were inactive in student moral reasoning. It makes sense to make use of developmental components that *are active* in order to stimulate *inactive* components. For instance, in the case of the Decliners in this study, educators could appeal to students' sense of personal agency and desire to engage in action for the benefit of others (both found to be fairly active in Decliners' writing) while simultaneously challenging the inconsistency of wanting to

help only those within our in-groups. Since this course does not include specific extracurricular, experiential or service components (see discussion above), creating opportunities and experiences appropriate to this course's format to help students encounter the legitimate needs and demands of "the Other" are important antidotes for personal interest, ego-centric and dualist tendencies. Noting the preference that *Gainers* showed for the juxtaposition of theory with figures and thinkers admired for the sacrifices they made for their principles suggests that using historical and current exemplars may help students understand what Postconventional thinking actually looks like. Including current examples of this sort of excellence is important since students at lower levels seem to use powers of abstraction to dismiss as "things that happened in the past" the pertinent challenges to their egocentric and in-group thinking.

Finally, in light of the results of this study, one might argue that convening students in homogenous groups for course activities, discussions and assignments would be advisable so that strategies tailored to various developmental tasks could be employed. Some research into best practices related to students at different stages of development suggest that students at varying levels of development need different and at times even conflicting modes of pedagogical support. Knefelkamp and Widick's 1974 work with Perry's scale of intellectual and ethical development suggest that a fruitful challenge for students at late multiplicity and contextual relativist stages would include low degrees of structure within instruction (Stephenson & Hunt, 1977). These researchers also found that a high degree of structure is an important support mode within instruction for student at dualist and early multiplicity stages. However, combined research on the usefulness of presenting competing perspectives and the positive effects of diverse peer interaction

suggests that this strategy may backfire and leave students to simply reinforce their own points of view. Indeed, many moral development researchers surmise that it is the experience of heterogeneity within the college experience (encountering new ideas, new people from different backgrounds and with different beliefs and values) itself that is at the heart of the positive effects of college participation on student moral development (King and Mayhew, 2002, 2008; Mayhew, Seifert & Pascarella, 2010, 2012; McNeel, 1991, 1994; Mayhew, Wolniak & Pascarella, 2008; Maeda, Thoma, & Bebeau, 2009; Pascarella, Seifert & Blaich, 2010; Pascarella & Terenzini 2005; Narvaez, Bebeau & Thoma, 1999b).

8.4.A Effective Practices of the Course

Many of the practices recommended by moral reasoning research are featured in the course investigated in this study. A thorough apprehension of the results of this study and of moral reasoning development literature confirm that many practices embedded in liberal education formats like this one are quite impactful in advancing moral reasoning among students. In particular, the qualitative findings of this study revealed that specific course features were especially effective in prompting distinct aspects of moral reasoning development. Thus, it appears to be true that the course's wide array of effective procedures and pedagogical methods provided enough breadth to reach students across a range of development. Results of the study show that certain course features were particularly notable.

First, the course's year-long format added greatly to its effectiveness in engaging students in a sustained, dialogical and communal approach to questions of justice and ethics. Particularly notable was that as a 12 credit course, it represented a substantial

percentage (on average 40%) of a student's first year GPA and as such was able to command students' attention to its themes and coursework. Meeting for 6 hours weekly in small class settings (all sections of the course are capped at 25 students) established an environment well suited for thorough and ongoing discussions about the "big questions" of life. Moreover, the course's restriction to freshmen only allowed for more comfortable class participation and allowed students at similar, though not homogeneous, levels of moral development to approach and work through these challenging themes and questions together.

Though extracurricular experiential learning components that have been found to be associated with moral growth, such as service learning (Gorman, Duffy, Heffernan, 1994), were not part of this course, experiences of sustained conversations, movie and documentary film viewings, written reflections and assigned attendance at campus lectures were regular features of the course that supplemented course texts. Faculty in the program are encouraged to engage students in discussion formats and many employ Socratic modes of discussion, which emphasizes not only the centrality of questions but also seeks to bring interlocutors to points of interruption and logical crisis wherein limitations of lower order thinking are exposed and reconsidered. Thus, coursework, course activities and the course format are designed to interrupt lower order moral thinking by not only exposing students to ideas of justice and ethics, but by engaging in sustained philosophical and theological attempts to resolve deep problems in these areas. Qualitative findings from this study show that these practices were associated with the mobilization of a variety of facets of student moral reasoning. Students who exhibited high levels of development consistently noted the impact of ethical and ethical-political

theory and the connection of theory with real world concerns on their ability to apprehend and process increasingly complex aspects of living a “good life.” Students who showed great gains demonstrated in pre-tests their readiness for bringing previous assumptions into question, supporting very recent research into important distinctions between moral transition and consolidation periods in student moral development (Mayhew, Seifert & Pascarella, 2012).

What is less clear is why these aspects of the course did not keep some students from losing ground in their moral reasoning development. Results from the study found that basic levels of anxiety, evidenced in Decliner tendency to emphasize the need for balance and safety in both pre- and post-essays, combined with the backward-looking orientation of most Decliner essays suggest that lower order thinking and habituation may be associated with moral and ethical retrenchment that resists these educational strategies. Thus, new and creative strategies are always welcome and should be actively sought, particularly with these students in mind.

8.5 IMPLICATIONS FOR LIBERAL EDUCATION PROGRAM DESIGN AND IMPLEMENTATION

Educators involved in the design and implementation of liberal education courses and programs do not need to be convinced that they should actively attend to the intellectual development of their students. They do this all the time, methodically and responsibly, and they find ways to regularly assess if their students are meeting their expectations. But when it comes to other developmental advances that liberal education espouses and claims to effect, educators are often at a loss as to how they might methodically and responsibly foster this pursuit. Furthermore, moral development

assessment is time consuming and difficult since its components are imprecise and hard to quantify. Too often, we rely on anecdotal evidence and hope for the best. Finally, there are important post-modern and contemporary criticisms of education's ability to promote moral growth and general ambivalence about the educator's role in that development. However, it remains the case that: 1) research shows clearly that college students are experiencing great gains in moral reasoning; 2) college participation is linked to advanced moral reasoning; and 3) the liberal education paradigm claims moral development as a hallmark objective; and 4) liberal education paradigm has been shown in research to be a more effective educational paradigm in promoting moral reasoning development than other educational contexts (King and Mayhew, 2002, 2008; Mayhew, Seifert & Pascarella, 2010, 2012; Mentkowski, 2000; Pascarella, 1997; Pascarella, Seifert & Blaich, 2010; Pascarella & Terenzini, 2005; Pascarella, Wolniak, Seifert & Blaich, 2005). With these four points in mind and in light of liberal education's need to defend its place in the American higher education scene, it only makes sense that those of us who work in liberal education begin to more accurately and robustly evaluate our work in this area of development. The answer to the question, *can liberal arts education be assessed?* is that it can and should be assessed. Liberal education *does* make a difference in the moral development of students, as demonstrated in this investigation of a sustained, carefully designed liberal arts curriculum at this institution. Uncovering the extent to which these results may be generalized to other liberal education programs, courses and contexts demands much further analysis and assessment. That will entail a great deal of effort, resources and willingness on the part of liberal education professionals to develop

tools that suit the task. A “one-size fits all” approach will not suffice for the many aspects of student development that we seek to explore.

Finding and developing adequate tools to measure this wide array of developmental trajectories are central of this endeavor. Familiarizing educators with developmental theory and working with a variety of accepted, theory-based measures of development are first steps toward gaining a real view of what student development looks like and how we can most effectively promote it. This study suggests that other important steps are using mixed-methods approaches to assessment, since in this case, analysis of student writing illuminated aspects of student development and engagement in coursework that were not readily apparent in quantitative analysis. Multi-faceted investigations may help us ascertain not only advancements in moral reasoning but also help distinguish the specific conditions needed to spur and sustain development. Analysis of writing is only one option among many for allowing student perception to add breadth to our evaluation processes. Interviewing students may also be a suitable option for this type of analysis. Both formats complement a measure like the DIT, which utilizes a *recognition* model of assessment that is aptly supplemented by a *production* format of student writing or interviews (Rest et al., 1999b).

The variety of developmental levels and gains found in this study further suggest that educators would profit from conceiving of a variety of milestones in moral reasoning advancement to more adequately capture the essential features of development. Just as we would not recommend one-size fits all assessment formats, neither should we use a one-size fits all approach to construing developmental objectives for our students. By having a better grasp of how students enter, change in and leave our programs, educators

would have a better sense of what might be reasonably expected from different students and would be able to better interpret when, how and to what extent gains are achieved.

Assessment tools like the one at the center of this study seek to reveal what can be known about the development of student moral reasoning and how effective our educational programs are in promoting this development. Understanding more about how students approach, perceive, work out and make meaning out of moral and ethical questions will enable us to be more effective in designing and implementing impactful coursework and pedagogy to that end. The assessment format of this study was successful but it can be improved upon. Identifying specific variables within coursework, course activities, classroom practices and classroom environments would add greatly to a study of this sort. This assessment can and should be developed and replicated as either continued, program-wide assessment or as individual class assessments within similar types of courses.

The results of this study are suggestive of further areas of research that would add substantially to our understanding of student moral reasoning development and moral development more generally. Longitudinal studies would allow us to discover if the gains made by students are long lasting and upwardly directed. Work in this area would also provide a better sense of how first-year college effects come to bear on development in the rest of college and beyond. A revised version of the present study might seek to identify why groups of students at different developmental levels are better or worse at receiving and processing particular course material and activities and how those differences impact students' capacity for growth. Finally, it is important to remember that course and program assessment primarily seeks improved ways of helping students

flourish. This means that assessment ought to always have a particular eye to creating and developing our pedagogy, our course materials and our interactions with students. If our students are not better off when they leave our classes, intellectually, morally and personally, then liberal education is not doing its job well. We may not yet know *the best ways* to challenge and support student moral reasoning development, but as we improve our assessment tools and pay closer attention to this often disregarded developmental area, we may know *better ways* to offer the challenges and support students need.

References

- Appleyard, Joseph, (2008) *The Journey into Adulthood: Understanding Student Formation*. Chestnut Hill, MA: Boston College Office of Marketing Communications. Retrieved from <http://www.bc.edu/content/dam/files/offices/mission/pdf1/umm1.pdf>
- Armon, C., & Dawson, T. L. (1997). Developmental trajectories in moral reasoning across the life span. *Journal of Moral Education*, 26(4), 433.
- Arum, R., & Roksa, J. (2010). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.
- Association of American Colleges and Universities (AAC&U). (2002). *Great expectations: A new vision of learning as a nation goes to college*. Washington, DC: AAC&U.
- Astin, A. (1993). *What matters in college? four critical years revisited*. San Francisco, CA: Jossey Bass.
- Baker, V. L., Baldwin, R. G., & Makker, S. (2012). Where are they now? revisiting breneman's study of liberal arts colleges. *Liberal Education*, 98(3), 48-53.
- Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, 31(2), 101-119.
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, 71(2), 364.
- Baumrind, D. (1986). Sex differences in moral reasoning: Response to Walker's (1984) conclusion that there are none. *Child Development*, 57(2), 511-581.

- Bebeau, M. (1994). Influencing the moral dimensions of dental practice. In Rest, James & Narváez, D. (Eds.), *Moral development in the professions: Psychology and applied ethics* (121-146). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bebeau, M. (2001). Influencing the moral dimensions of professional practice: Implications for teaching and assessing for research integrity. *Proceedings of the ORI Conference on Research on Research Integrity*. Retrieved from www.personal.umich.edu/~nsteneck/rcri/index.html.
- Bebeau, M. (2002). The defining issues test and the four component model: Contributions to professional education. *Journal of Moral Education*, 31(3), 271-95.
- Bebeau, M., & Brabeck, M. (1987). Integrating care and justice issues in professional moral education: A gender perspective. *Journal of Moral Education*, 16(3), 189-203.
- Bebeau, M. J., & Thoma, S. J. (1994). The impact of a dental ethics curriculum on moral reasoning. *Journal of Dental Education*, 58(9), 684-692.
- Bebeau, M. J., & Thoma, S. J. (1999). 'Intermediate' concepts and the connection to moral education. *Educational Psychology Review*, 11(4), 343.
- Berrett, D. (2012, September 18, 2012). All about the money. *Chronicle of Higher Education*.
- Berube, C. T., & Berube, M. R. (2010). *The moral university*. Lanham, Md.: Rowman & Littlefield Publishers.
- Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, 88(1), 1-45.

- Bok, D. (2007). *Our underachieving colleges: A candid look at how much students are learning and why they should be learning more*. Princeton, NJ: Princeton University Press.
- Boss, J. A. (1994). The anatomy of moral intelligence. *Educational Theory*, 44(4), 399-416.
- Boston College Institutional Research, Planning & Assessment. (2013). *NSSE survey summary report*. Chestnut Hill, MA: Boston College Institutional Research, Planning & Assessment.
- Bowen, W. G., Kurzweil, M. A., Tobin, E. M., & Pichler, S. C. (2005). *Equity and excellence in american higher education*. Charlottesville: University of Virginia Press.
- Bowman, N. A., & Seifert, T. A. (2011). Can college students accurately assess what affects their learning and development? *Journal of College Student Development*, 52(3), 270-290.
- Brabeck, M. (1983). Moral judgment: Theory and research on differences between males and females. *Developmental Review*, 3(3), 274-291.
- Brabeck, M. (1984). Longitudinal studies of intellectual development during adulthood: Theoretical and research models. *Journal of Research and Development in Education*, 17(3), 12-27.
- Brubacher, J. S., & Rudy, W. (1997). *Higher education in transition: A history of american colleges and universities* (Fourth ed.). New Brunswick, NJ: Transaction Publishers.

- Bruess, B. J. (1995). A cross-sectional study examining the moral reasoning and identity development of first-year and graduating college students. (Unpublished doctoral dissertation). Ohio University, Athens, OH.
- Bruess, B. J., & Pearson, F. C. (2002). The debate continues: Are there gender differences in moral reasoning as defined by Kohlberg? *College Student Affairs Journal*, 21(2), 38-52.
- Burwell, R., Butman, R., & Van Wicklin, J. (1992). *Values assessment at three consortium colleges: A follow-up study*. (Unpublished doctoral dissertation). Houghton College, Houghton, NY.
- Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48(3), 306-307.
- Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 15(2), 195-207.
- Carlson, S. (2013, April 22). How to assess the real payoff of a college degree. *Chronicle of Higher Education*.
- Colby, A. (1983). A longitudinal study of moral judgment. *Monographs of the Society for Research in Child Development*, v48 n1-2 p1-124 1983, 48(1-2)
- Colby, A., & Ehrlich, T. (2000). *Civic Responsibility and Higher Education*. New York, NY: Rowman & Littlefield Press.
- Colby, A., Ehrlich, T., Beaumont, E., & Stephens, J. (2003). Educating undergraduates for responsible citizenship. *Change: The Magazine of Higher Learning*, 35(6), 40-48.

- Colby, A., Ehrlich, T., Sullivan, W. M., & Dolle, J. R. (2011). *Rethinking undergraduate business education: Liberal learning for the profession*. San Francisco, CA: Jossey-Bass.
- Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment*. Cambridge: Cambridge University Press.
- Constantinople, A. (1969). An eriksonian measure of personality development in college students. *Developmental Psychology*, 1, 357-372.
- Cooper, D. L., Liddell, D. L., Davis, T. J., & Pasquesi, K. (2012). Providing evidence in the moral domain. *New Directions for Student Services*, 2012(139), 93-103.
- Cox, H. G. (1985). Moral reasoning and the humanities. *Liberal Education*, 71(3), 195-203.
- Creswell, J., Clark, J., Gutmann, V., Hanson, W., Clark, V., & Creswell, J. (2008). An expanded typology for classifying mixed methods research into designs. *The Mixed Methods Reader*, 159-196.
- Dalton, J. C., Russell, T. R., & Kline, S., (2004). *Assessing character outcomes in college*. San Francisco: Jossey-Bass.
- Davey, J. D. (2001). *The conscience of the campus: Case studies in moral reasoning among today's college students*. Westport, CT: Praeger.
- Davison, M. L. (1977). On a metric, unidimensional unfolding model for attitudinal and developmental data. *Psychometrika*, 42(4), 523-548.
- Davison, M. L., & and Others. (1981). *Higher education and cognitive-social development project. final report. Project report no. 8, Report: ED208771*.

- Delbanco, A. (2012). *College: What it was, is, and should be*. Princeton, NJ: Princeton University Press.
- Derry, S. J. (1996). Cognitive schema theory in the constructivist debate. *Educational Psychologist, 31*(3-4), 163-174.
- Elm, D. R., & Weber, J. (1994). Measuring moral judgment: The moral judgment interview or the defining issues test? *Journal of Business Ethics, 13*(5), pp. 341-355.
- Emler, N., Palmer-Canton, E., & James, A. (1998). Politics, moral reasoning and the defining issues test: A reply to Barnett et al. *British Journal of Social Psychology, 37*(4), 457-476.
- Emler, N., Renwick, S., & Malone, B. (1983). The relationship between moral reasoning and political orientation. *Journal of Personality and Social Psychology, 45*(5), 1073.
- Engberg, M. E., & Hurtado, S. (2011). Developing pluralistic skills and dispositions in college: Examining racial/ethnic group differences. *The Journal of Higher Education, 82*(4), 416-443.
- Engberg, M. E., Hurtado, S., & Smith, G. C. (2007). Developing attitudes of acceptance toward lesbian, gay, and bisexual peers: Enlightenment, contact, and the college experience. *Journal of Gay & Lesbian Issues in Education, 4*(3), 49-77.
- Engberg, M. E., & Mayhew, M. J. (2007). The influence of first-year “success” courses on student learning and democratic outcomes. *Journal of College Student Development, 48*(3), 241-258.
- Engberg, M. E., & Porter, K. B. (2013). Conceptualizing the linkages between diversity experiences and moral development. *Journal of College and Character, 14*(4), 289-300.

- Evens, J. M. (1995). Indexing moral judgment using multidimensional scaling. (Unpublished doctoral dissertation) University of Minnesota, Minneapolis, MN.
- Finger, W. W., Borduin, C. M., & Baumstark, K. E. (1992). Correlates of moral judgment development in college students. *The Journal of Genetic Psychology, 153*(2).
- Fish, S. (2003). Aim low. *Chronicle of Higher Education, 49*(36), C5.
- Fish, S. (2008). *Save the world on your own time*. New York, NY: Oxford University Press.
- Fish, S. (2010). I know it when I see it: A reply to Kiss and Euben. In E. Kiss, & J. P. Euben (Eds.). *Debating moral education: Rethinking the role of the modern university* (76-91). Durham, NC: Duke University Press.
- Friedman, A. A. (1995). The relationship between intellectual disposition and reflective judgment in college women. (Unpublished doctoral dissertation). Boston College, Chestnut Hill, MA.
- Frimer, J. A., Biesanz, J. C., Walker, L. J., & MacKinlay, C. W. (2013). Liberals and conservatives rely on common moral foundations when making moral judgments about influential people. *Journal of Personality and Social Psychology, 104*(6), 1040-1059.
- Garver, E. (1985). Teaching writing and teaching virtue. *Journal of Business Communication, 22*(1), 51-73.
- Getz, I. R. (1984). Moral judgment and religion: A review of the literature. *Counseling and Values, 28*(3), 94-116.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

- Gorman, M., & Duffy, J. (1994). Service experience and the moral development of college students. *Religious Education*, 89(3), 422.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23(2), 101-124.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274.
- Grunwald, H. E., & Mayhew, M. J. (2008). Using propensity scores for estimating causal effects: A study in the development of moral reasoning. *Research in Higher Education*, 49(8), 758-775.
- Gutmann, A. (1999). *Democratic education* (2nd ed.). Princeton: Princeton University Press.
- Halliday, R., & Frantis, L. (2006). Teaching goodness: Moral development theory and the teaching of ethics. *Teaching Philosophy*, 29(2), 81-91.
- Hauerwas, S. (2010). The pathos of the university: The case of Stanley Fish. In E. Kiss, & J. P. Euben (Eds.). *Debating moral education: Rethinking the role of the modern university* (92-110). Durham, NC: Duke University Press.
- Hawkins, P. E. (2003). The development and conservation of moral reasoning at three liberal arts colleges. (Unpublished doctoral dissertation) Georgia State University, Atlanta, GA.
- Herington, C., & Weaven, S. (2008). Improving consistency for DIT results using cluster analysis. *Journal of Business Ethics*, 80(3), pp. 499-514.

- Hersh, R. H., & Schneider, C. G. (2005). Fostering personal & social responsibility on college & university campuses. *Liberal Education*, 91(3), 6-13.
- Higher education act amendments of 1998: Report of the committee on labor and human resources, together with additional views to accompany S. 1882*(1998). Washington: U.S. G.P.O.
- Hirst, P. H. (1965). Liberal education and the nature of knowledge. *Philosophical Analysis and Education*, 2, 113-140.
- Hoekema, D. (2010). Is there an ethicist in the house? how can we tell? In E. Kiss, & J. P. Euben (Eds.). *Debating moral education: Rethinking the role of the modern university* (249-266). Durham, NC: Duke University Press.
- Hoffman, M. L. (2001). *Empathy and moral development: Implications for caring and justice*. Cambridge: Cambridge University Press.
- Houck, E. C. (2010). The relationship between moral reasoning and the multidimensional model of racial identity. (Unpublished doctoral dissertation). Tennessee State University, Nashville, TN.
- Hurtado, S. (2007). Linking diversity with the educational and civic missions of higher education. *The Review of Higher Education*, 30(2), 185-196.
- Hurtado, S., Mayhew, M. J., & Engberg, M. E. (2012). Diversity courses and students' moral reasoning: A model of predispositions and change. *Journal of Moral Education* 41(2), 201-224.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 602-611.

- Kay, S. R. (1982). Kohlberg's theory of moral development: Critical analysis of validation studies with the defining issues test. *International Journal of Psychology*, 17(1), 27-42.
- Keohane, N. O. (2006). *Higher ground: Ethics and leadership in the modern university*. Durham, NC: Duke University Press.
- Kerr, C. (1963). *The uses of the university* (Fourth ed.). Cambridge, MA: Harvard University Press.
- King, P. M. (1997). Character and civic education: What does it take? *Educational Record*, 78(3-4), 87-93.
- King, P. M. (2009). Principles of development and developmental change underlying theories of cognitive and moral development. *Journal of College Student Development*, 50(6), 597-620.
- King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. *Jossey-Bass higher and adult education series and jossey-bass social and behavioral science series*. San Francisco: Jossey-Bass.
- King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist*, 39(1), 5-18.
- King, P. M., & Mayhew, M. J. (2002). Moral judgement development in higher education: Insights from the defining issues test. *Journal of Moral Education*, 31(3), 247-270.

- Kiss, E., & Euben, J. P. (Eds.). (2010). *Debating moral education: Rethinking the role of the modern university*. Durham, NC: Duke University Press.
- Kiss, E., & Euben, J. P. (2010). Debating moral education: An introduction. In E. Kiss, & J. P. Euben (Eds.). *Debating moral education: Rethinking the role of the modern university* (3-25). Durham, NC: Duke University Press.
- Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice*. New York, NY: Harper & Row.
- Kohlberg, L. (1984). *The psychology of moral development: The nature and validity of moral stages*. New York, NY: Harper & Row.
- Kronman, A. (2007). *Education's end: Why our colleges and universities have given UP on the meaning of life*. New Haven, CT: Yale University Press.
- Lawrence, J. A. (1987). Verbal processing of the defining issues test by principled and non-principled moral reasoners. *Journal of Moral Education*, 16(2), 117-130.
- Lefond, D. C. (1979). The assessment of moral reasoning, sex, class status and group norms as predictors of college student moral behavior. (Unpublished doctoral dissertation). Case Western Reserve University, Cleveland, OH.
- Lewin, T. (2013, October 30). As interest fades in the humanities, colleges worry. *New York Times*.
- Lewis, H. (2006). *Excellence without A soul: Does liberal education have a future?* New York, NY: Perseus Books.
- Liddell, D. L. (1998). Comparison of semistructured interviews with a quantitative measure of moral orientation. *Journal of College Student Development*, 39(2), 169-178.

- Lifton, P. D. (1985). Individual differences in moral development: The relation of sex, gender, and personality to morality. *Journal of Personality*, 53(2), 306-334.
- Lind, G. (2010). The meaning and measurement of moral judgment competence: A dual-aspect model. In Daniel Fasko (Ed.), *Contemporary philosophical and psychological perspectives on moral development and education*. Cresskill, NJ : Hampton Press, 185-220.
- Lind, G., Hartmann, H., & Wakenhut, R. (Eds.). (1985). *Moral development and the social environment*. Chicago: Precedents Publishing.
- Linstrum, K. S. (2009). Ethical training, moral development, and ethical decision making in master's-level counseling students. *Journal of College and Character*, 10(3).
- Loescher, K. J. (2004). The impact of an "ethics across the curriculum" initiative on the cognitive moral development of business school undergraduates. (Unpublished doctoral dissertation). Nova Southeastern University, Fort Lauderdale, FL.
- Maeda, Y., Thoma, S. J., & Bebeau, M. J. (2009). Understanding the relationship between moral judgment development and individual characteristics: The role of educational contexts. *Journal of Educational Psychology*, 101(1), 233.
- Marsden, G. M. (1994). *The soul of the american university: From protestant establishment to established nonbelief*. Oxford: Oxford University Press.
- Maxwell, B. (2010). Does ethical theory have a place in post-Kohlbergian moral psychology? *Educational Theory*, 60(2), 167-188.
- Mayhew, M. J. (2004). Curricular content and pedagogical practices that influence the development of moral reasoning in undergraduate students. (Unpublished doctoral dissertation). University of Michigan, Ann Arbor, MI.

- Mayhew, M. J. (2012). A multilevel examination of the influence of institutional type on the moral reasoning development of first-year students. *Journal of Higher Education*, 83(3), 367-388.
- Mayhew, M. J., & Engberg, M. E. (2010). Diversity and moral reasoning: How negative diverse peer interactions affect the development of moral reasoning in undergraduate students. *The Journal of Higher Education*, 81(4), 459-488.
- Mayhew, M. J., & Fernandez, S. D. (2007). Pedagogical practices that contribute to social justice outcomes. *Review of Higher Education*, 31(1), 55-80.
- Mayhew, M. J., & King, P. (2008). How curricular content and pedagogical strategies affect moral reasoning development in college students. *Journal of Moral Education*, 37(1), 17-40.
- Mayhew, M. J., Seifert, T. A., & Pascarella, E. T. (2010). A multi-institutional assessment of moral reasoning development among first-year students. *Review of Higher Education*, 33(3), 357-390.
- Mayhew, M. J., Seifert, T. A., & Pascarella, E. T. (2012). How the first year of college influences moral reasoning development for students in moral consolidation and moral transition. *Journal of College Student Development*, 53(1), 19-40.
- Mayhew, M. J., Vanderlinden, K., & Kim, E. K. (2010). A multi-level assessment of the impact of orientation programs on student learning. *Research in Higher Education*, 51(4), 320-345.
- Mayhew, M. J., Wolniak, G. C., & Pascarella, E. T. (2008). How educational practices affect the development of life-long learning orientations in traditionally-aged undergraduate students. *Research in Higher Education*, 49(4), 337-356.

- McCarthy, M. A., Phillips, B. A., Mills, J., & Horn, M. E. (2002). Moral reasoning: Does the college experience make a difference? *College Student Affairs Journal*, 21(2), 3-8.
- McGuire, P. P. (1994). A theology class: A means of fostering intellectual and moral change in college students. (Unpublished doctoral dissertation). Columbia University Teachers College, New York, NY.
- McNeel, S. P. (1991). Christian liberal arts education and growth in moral judgment. *Journal of Psychology and Christianity*, 10(4), 311.
- McNeel, S. P. (1994). College teaching and student moral development. *Moral development in the professions: Psychology and applied ethics* (27-49). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mealey, R. M. (2004). *The relationship between moral development and moral action: College students and organ donation*. (Unpublished doctoral dissertation): Boston College, Chestnut Hill, MA.
- Mentkowski, M., & Strait, M. J. (1983). *A longitudinal study of student change in cognitive development, learning styles, and generic abilities in an outcome-centered liberal arts curriculum: Final report to the national institute of education: Research report number 6*. Milwaukee, WI: Alverno College Productions.
- Mentkowski, M., Much, N. C., & Giencke-Holl, L. (1983). *Careering after college: Perspectives on lifelong learning and career development: Final report to the national institute of education: Research report number 8*. Milwaukee, WI: Alverno College Productions.

- Mentkowski, M. (1991). *Understanding abilities, learning and development through college outcomes studies: What can we expect from higher education assessment?* Milwaukee, WI: Office of Research and Evaluation, Alverno College.
- Moreland, C., & Leach, M. M. (2001). The relationship between black racial identity and moral development. *Journal of Black Psychology*, 27(3), 255-271.
- Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40(2), 120-123.
- Mulcahy, D. G. (2008). Newman's theory of a liberal education: A reassessment and its implications. *Journal of Philosophy of Education*, 42(2), 219-231.
- Murk, D. A., & Addleman, J. A. (1992). Relations among moral reasoning, locus of control, and demographic variables among college students. *Psychological Reports*, 70(2), 467-476.
- Mustapha, S. L., & Seybert, J. A. (1991). Moral reasoning in college students: Effects of two general education curricula. *Educational Research Quarterly*, 14(4), 32-40.
- Narvaez, D. (1998). The influence of moral schemas on the reconstruction of moral narratives in eighth graders and college students. *Journal of Educational Psychology*, 90(1), 13-24.
- Narvaez, D. (1999). Using discourse processing methods to study moral thinking. *Educational Psychology Review*, 11(4), 377-393.
- Narvaez, D. (2001). Moral text comprehension: Implications for education and research. *Journal of Moral Education*, 30(1), 43-54.

- Narvaez, D. (2008). Human flourishing and moral development: Cognitive and neurobiological perspectives of virtue development. *Handbook of Moral and Character Education*, New York, NY: Routledge.
- Narvaez, D., & Bock, T. (2002). Moral schemas and tacit judgement or how the defining issues test is supported by cognitive science. *Journal of Moral Education*, 31(3), 297-314.
- Narvaez, D., Getz, I., Rest, J. R., & Thoma, S. J. (1999). Individual moral judgment and cultural ideologies. *Developmental Psychology*, 35(2), 478-488.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231.
- Nucci, L. & Pascarella, E. (1987). The influence of college on moral development. In J. Smart (Ed.), *Higher education: Handbook of theory and research*, vol. 3 (271-326). New York, NY: Agathon Press.
- Nunner-Winkler, G. (1984). Two moralities? A critical discussion of an ethic of care and responsibility versus an ethic of rights and justice. *Morality, Moral Behavior, and Moral Development*, 348-361.
- Nussbaum, M. (2004). Liberal education and global community. *Liberal Education*, 90(1), 42-47.
- Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools*, 13(1)
- Pascarella, E. T. (1997). College's influence on principled moral reasoning. *Educational Record*, 78(3-4), 47-55.

- Pascarella, E. T. (2005). *Liberal arts colleges and liberal arts education: New evidence on impacts*. San Francisco, CA: Wiley.
- Pascarella, E. T. (2005). *Cognitive impacts of the first year of college*. Mahwah: Lawrence Erlbaum Associates.
- Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How robust are the findings of “academically adrift”? *Change: The Magazine of Higher Learning*, 43(3), 20-24.
- Pascarella, E. T., Seifert, T. A., & Blaich, C. (2010). How effective are the NSSE benchmarks in predicting important educational outcomes? *Change: The Magazine of Higher Learning*, 42(1), 16-22.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco, CA: Jossey-Bass.
- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: Volume 2, a third decade of research*. San Francisco, CA: Jossey-Bass.
- Pascarella, E., Wolniak, G., Seifert, T., Cruce, T., & Blaich, C. (2005). *Liberal arts colleges and liberal arts education: New evidence on impacts: ASHE higher education report (J-B ASHE higher education report series (AEHE))*. San Francisco, CA: Jossey-Bass.
- Piaget, J. (1997). *The moral judgement of the child*. New York, NY: Simon and Schuster.
- Pearson, F. C., & Bruess, B. J. (2001). *The gender debate about identity and moral development continues: What about the men?* Paper presented at the Annual Meeting of the National Association of Student Personnel Administrators. Seattle, WA, March 17-21, 2001. Retrieved from <http://files.eric.ed.gov/fulltext/ED455177.pdf>.

- Penn, W. Y., Jr. (1990). Teaching ethics: A direct approach. *Journal of Moral Education*, 19(2), 124-138.
- Penny, N. (2008). Assessing student outcomes of a values-based education: The effectiveness of the defining issues test. (Unpublished doctoral dissertation). Wilmington University, Wilmington, DE.
- Perry, W. (1970). *Forms of intellectual and ethical development in the college years*. New York, NY: Holt, Rinehart & Winston.
- Perry, W. (1981). Cognitive and ethical growth: The making of meaning. In A. Chickering (Ed.), *The modern american college* (76-116). San Francisco: Jossey-Bass.
- Perry, W. (1999). *Forms of intellectual and ethical development in the college years: A scheme. jossey-bass higher and adult education series*. San Francisco, CA: Jossey-Bass Publishers.
- Plano Clark, V. L., & Creswell, J. W. (2008). *The mixed methods reader*. Thousand Oaks, CA: Sage.
- Price, B. R. (1992). Meta-analysis of the effects of college on moral reasoning. (Unpublished doctoral dissertation). Arizona State University, Phoenix, AZ.
- Rawls, J. (2001). *Justice as fairness*. Cambridge, MA: Harvard University Press.
- Reed, D. C. (2008). A model of moral stages. *Journal of Moral Education*, 37(3), 357-376.
- Reiman, A. (2002). A comparison of four year longitudinal studies of postconventional moral judgment reasoning in teacher education and other selected undergraduate samples. *American Educational Research Association*, New Orleans, LA.

- Reiman, A. J. (2004). Longitudinal studies of teacher education candidates' moral reasoning and related promising interventions. *Journal of Research in Character Education, 2*(2), 141-150.
- Rest, J. R. (1979). *Development in judging moral issues*. Minneapolis, MN: University of Minnesota Press.
- Rest, J. R. (1985). Moral judgment development and formal education. *Developmental Psychology, 21*, 709-714.
- Rest, J. R. (1987). *Guide for the defining issues test*. Minneapolis, MN: University of Minnesota Press.
- Rest, J. R., Narvaez, D., Bebeau, M., & Thoma, S. (1999). A neo-kohlbergian approach: The DIT and schema theory. *Educational Psychology Review, 11*(4), 291-324.
- Rest, J. R., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (1999). *Postconventional moral thinking: A neo-Kohlbergian approach*. Mahwah, NJ: L. Erlbaum Associates.
- Rest, J. R., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (1999). DIT2: Devising and testing a revised instrument of moral judgment. *Journal of Educational Psychology, 91*(4), 644-659.
- Rest, J. R., Narvaez, D., Thoma, S., & Bebeau, M. (2000). A neo-Kohlbergian approach to morality research. *Journal of Moral Education, 29*(4), 381-395.
- Rest, J. R., & Thoma, S. J. (1985). Relation of moral judgment development to formal education. *Developmental Psychology, 21*(4), 709-714.
- Rest, J., Thoma, S., & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preference and stage consistency approaches. *Journal of Educational Psychology, 89*(1), 5-28.

- Rest, J. R., Thoma, S. J., Narvaez, D., & Bebeau, M. J. (1997). Alchemy and beyond: Indexing the defining issues test. *Journal of Educational Psychology*, 89(3), 498.
- Reuben, J. (1996). *The making of the modern university: Intellectual transformation and the marginalization of morality*. Chicago: University of Chicago Press.
- Reuben, J. (2010). The changing contours of moral education in american colleges and universities. In E. Kiss, & J. P. Euben (Eds.). *Debating moral education: Rethinking the role of the modern university* (27-55). Durham, NC: Duke University Press.
- Reynolds, S. J., & Ceranic, T. L. (2007). The effects of moral judgment and moral identity on moral behavior: An empirical examination of the moral individual. *Journal of Applied Psychology*, 92(6), 1610.
- Rogers, G. (2002). Rethinking moral growth in college and beyond. *Journal of Moral Education*, 31(3), 325-338.
- Saldaña, J. (2011). *Fundamentals of qualitative research*. New York: Oxford University Press.
- Schlaefli, A., Rest, J. R., & Thoma, S. J. (1985). Does moral education improve moral judgment? A meta-analysis of intervention studies using the defining issues test. *Review of Educational Research*, 55(3), pp. 319-352.
- Scott, J. (2006). The mission of the university. *Journal of Higher Education*, 77(1), 2-39.
- Shaub, M. (1994). An analysis of factors affecting the cognitive moral development of auditors and auditing students. *Journal of Accounting Education*, 12(1), 1-26.
- Shear, M. D. (2014). Colleges rattled as obama seeks rating system. *New York Times*, may, 26, A1.

- Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, 37(3), 330-349.
- Skoe, E. E. A. (2010). The relationship between empathy-related constructs and care-based moral development in young adulthood. *Journal of Moral Education*, 39(2), 191-211.
- Sloan, D. (1994). *Faith and knowledge: Mainline protestantism and american higher education* Westminster: John Knox Press.
- Smith, C., Christoffersen, K. M., Davidson, H., & Herzog, P. S. (2011). *Lost in transition: The dark side of emerging adulthood*. New York: Oxford University Press.
- Snarey, J., & Samuelson, P. (2008). Moral education in the cognitive developmental tradition: Lawrence Kohlberg's revolutionary ideas. *Handbook of Moral and Character Education*. 53-79.
- Sottile, J. M., Jr., & Brozik, D. (2000). *Ethical behavior and human development among male and female graduate and undergraduate college students*. Retrieved from <http://eric.ed.gov.proxy.bc.edu/?id=ED443044>.
- Stake, R. E. (1995). *The art of case study research* New York: Sage.
- Stake, R. E. (2006). *Multiple case study analysis*. New York: The Guilford Press.
- Stephens, D. G. *A correlational study on parental attachment and moral judgment competence of millennial generation college students*. (Unpublished doctoral dissertation). Lincoln, NE: University of Nebraska

- Stephenson, B. W., & Hunt, C. (1977). Intellectual and ethical development: A dualistic curriculum intervention for college students. *The Counseling Psychologist*, 6(4) 39-42.
- Strange, C. C. (2004). Measuring up: Defining and assessing outcomes of character in college. *New Directions for Institutional Research*, 2004(122), 25-36.
- Sullivan, E. (2011) Can Ethics Be Taught? *A Quasi-Experimental Study of the Impact of Class Size on the Cognitive Moral Reasoning of Freshmen Business Students*. (Unpublished doctoral dissertation). Boston College, Chestnut Hill, MA.
- Tennant, S. B. (1990). *Personal and moral development: A developmental curriculum intervention for liberal arts freshmen*. (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.
- Thoma, S. J. (1986). Estimating gender differences in the comprehension and preference of moral issues. *Developmental Review*, 6(2), 165-180.
- Thoma, S. J. (2002). An overview of the minnesota approach to research in moral development. *Journal of Moral Education*, 31(3), 225-245.
- Thoma, S. J., Narvaez, D., Rest, J., & Derryberry, P. (1999). Does moral judgment development reduce to political attitudes or verbal ability? evidence using the defining issues test. *Educational Psychology Review*, 11(4), 325-341.
- Thoma, S. J., & Rest, J. R. (1999). The relationship between moral decision making and patterns of consolidation and transition in moral judgment development. *Developmental Psychology*, 35(2), 323-334.

- Thoma, S. J., Rest, J. R., & Davison, M. L. (1991). Describing and testing a moderator of the moral judgment and action relationship. *Journal of Personality and Social Psychology*, 61(4), 659-669.
- Thoma, S., Barnett, R., Rest, J., & Narvaez, D. (1999). What does the DIT measure? *British Journal of Social Psychology*, 38(1), 103-111.
- Trull, T. E. (1990). The effect of the freshman experience on moral development: Cognitive or affective? (Unpublished doctoral dissertation). Memphis State University, Memphis, TN.
- Vine, I. (1996). The measure of morality-rethinking moral reasoning. *Journal of Moral Education*, 25(4), 455-466.
- Walker, L. J. (1984). Sex differences in the development of moral reasoning: A critical review. *Child Development*, 55(3), pp. 677-691.
- Walker, L. J. (1986). Sex differences in the development of moral reasoning: A rejoinder to baumrind. *Child Development*, 57(2), 522.
- Walker, L. J. (2002). The model and the measure: An appraisal of the minnesota approach to moral development. *Journal of Moral Education*, 31(3), 353-367.
- Waterman, A. (1982). Identity development from adolescence to adulthood: An extension of theory and a review of research. *Developmental Psychology*, 18, 341-358.
- Weaver Frost, L. (1997). The relationship between human values and moral reasoning as components of moral behavior. (Unpublished doctoral dissertation). University of Southern California, Los Angeles, CA.
- Whiteley, J. M., Bertin, & Berry. (1980). Research on the development of moral reasoning of college students. *New Directions for Higher Education*, (31), 35-44.

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, 120(2), 202-214.